Combining Single-Molecule with Ensemble Fluorescence Experiments to Disentangle the Interplay of Folding and Activity of a Group II Intron Ribozyme

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Susann Zelger-Paulus

von/aus

Stans NW/Deutschland

Promotionskommission

Prof. Dr. Roland K.O. Sigel (Vorsitz)
Prof. Dr. Benoît Kornman
Prof. Dr. Gilles Gasser
Prof. Dr. Michael Hengartner

Zürich, 2018

Dedicated to my parents Magdalena and Eberhard Paulus

Meinen Eltern gewidmet, Magdalena und Eberhard Paulus

Acknowledgements

My special thanks go to Prof. Roland K.O. Sigel for the opportunity to conduct my Ph.D. in his group. Thank you for believing in me and giving me the freedom to develop this project, you made the balance between the Ph.D. and my family possible. I would like to thank also my Ph.D. committee members Prof. Dr. Benoît, Prof. Dr. Gilles Gasser, and Prof. Dr. Micheal Hengartner.

This work would not have been possible without the help of many people, which I would like to thank at this point. Especially, I would like to thank Dr. Richard Börner for his support and advice during the whole time. Thank you for your continued interest and enthusiasm for this project, it was a great help to discuss with you all the challenging questions along the way. I am especially grateful to Besim Fazliji for his work as my BUSS- and Master student. It was a great pleasure to work with you. Thanks to Dr. Mélodie Hadzic for our collaboration on the encapsulation of vesicles and the endless exciting discussions about optimizations.

Many thanks go to Dr. Sofia Gallo, Dr. Silke Johannsen, and Maya Gulotti-Georgieva for their unremitting support, their friendship, and the constructive criticism, which was a great help and driving me forward.

Huge thanks to Dr. Richard Börner, Besim Fazliji, Dr. Sofia Gallo, Maya Gulotti-Georgieva, and Fabio Steffen for their efforts in reading this thesis and giving constructive criticism.

Besides, I would like to thank all the present and former members of the smFRET subgroup for the support and enthusiasm, which made a pleasant working atmosphere.

Finally, I would like to emphasize the great support of my family; Tom and Simon, thank you for all your love, patience, belief, and motivation during this exciting time, which encouraged me every day. My parents, Magdalena und Eberhard Paulus who motivated me from far and my mother in law, Matha Zelger for all this overtime in baby care.

Financial support by the UZH Forschungskredit (FK-13-095, FK-15-09), the graduate school (CMSZH), and the COST – Action CM1105 is gratefully acknowledged.

Abstract

"Seeing is believing": imaging biomolecules is a fascinating and common technique to track, e.g. binding events or folding pathways. The focus of the present thesis lies on the visualization of the splicing process of a catalytic RNA that takes place during RNA maturation.

Self-splicing group II introns are fascinating examples of functional RNAs where structure and function are inextricably linked with each other. We are investigating the complex RNA folding of the mitochondrial group II intron *Sc*.ai5γ from *Saccharomyces cerevisiae* and monitor its self-cleavage reaction by single-molecule Förster Resonance Energy Transfer (smFRET). For this purpose, we are fluorescently labeling the RNA at various positions; within the intron, at its flanking exons and cross-wise. While RNA splicing *in vivo* is promoted by proteins, self-excision *in vitro*, on the other hand, requires higher concentrations of salt and elevated temperature. We are encapsulating the group II intron into surface-immobilized vesicles and initiate the splicing process by increasing the temperature *in situ* on the microscope. In this way, we can follow the transition of an inactive partially folded ribozyme towards its cleavage competent fold on a single-molecule level.

Kurzfassung

"Sehen bedeutet Glauben": Biomoleküle sichtbar machen ist eine spannende und weitverbreitete Methode, z.B. um einen Bindungsvorgangs oder eine Faltung zu untersuchen. Die Visualisierung des Spleissens einer katalytisch aktiven RNA während der RNA-Reifung ist Fokus dieser Arbeit. Selbst-schneidende Gruppe II Introns sind hochinteressante Beispiele funktionaler RNAs bei denen Struktur und Funktion untrennbar miteinander verbunden sind. Wir untersuchen den komplexen Mechanismus der RNA-Faltung am Beispiel des Gruppe II Introns *Sc.*ai5γ, welches sich im mitochondrialen Genome von *Saccharomyces cerevisiae* befindet. Zusätzlich verfolgen wir das Spleissen mittels Einzelmolekülfluoreszenzspektroskopie (smFRET). Um dies zu ermöglichen, wurde die RNA an verschiedenen Stellen, im Intron und an den Exons, fluoreszent markiert. Obwohl RNA-Spleissen *in vivo* Hilfsproteine benötigt, wird das Spleissen *in vitro* mittels hohen Salzkonzentrationen und erhöhter Temperatur erreicht. Wir kapseln die RNA in Vesikel ein, welche ihrerseits an eine Oberfläche binden, und erhöhen die Temperatur *in situ*. Damit können wir den Übergang zwischen einem inaktiven partiell ungefalteten Ribozym zu einem katalytisch aktiven, gefalteten Ribozym auf Einzelmolekülebene untersuchen.

Table of Contents

Abbreviations

17/7	Substrate, consists of 17 nt of the 5’exons followed by the first 7 nt of the intron, named 17/7
ALEX	Alternating laser excitation
Bp	Base pair
DNA	Deoxyribonucleic acid
EMCCD	Electron multiplying charge-coupled device
En	Endonuclease
hnRNA	Heterogeneous nuclear RNA
IBS/EBS	Intron-binding site, exon-binding site
IEP	Intron-encoded protein
LPR	Long pass filter
MASH	Multifunctional analysis software for heterogeneous data
mRNA	Messanger RNA
Mss116	ATP-dependent RNA helicase
NAIM	Nucleotide analog interference mapping
Nt	Nucleotides
NTP	Nucleotide triphosphate
OD	Optical density
ORF	Open reading frame
OSS	Oxygen scavenger system
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PNA	Peptide nucleic acids
RIFE	RNA-induced fluorescence enhancement
RNA	Ribonucleic acid

RNP	Ribonucleoprotein
RT	Reverse transcriptase
sALEX	Stroboscopic ALEX
Sc.ai5γ	Group IIB intron, ai5γ from *Saccharomyces cerevisiae*
SER	Spliced-exon reopening
SM	Single-molecule
SMV	Single-molecule video
smFRET	Single-molecule Förster Resonance Energy Transfer
snRNA	Small nuclear RNA
SPR	Surface plasma resonance
TIRF	Total internal reflection fluorescence
TPSPC	Time-correlated single photon counting

Chemicals

APS	Ammonium persulfate
Cy3/Cy5	Cyanine 3/5
DMPC	1,2-dimyristoyl-*sn*-glycero-3-phosphocholine
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
LB	Luria broth medium
MOPS	(3-(N-morpholino)propanesulfonic acid)
PE	1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine-N-(cap biotinyl)
PEG	Polyethylene glycol
TEMED	Tetramethylethylenediamine
Tris	Tris(hydroxymethyl) aminomethane

1. Chapter

Motivation and research background

1.1 Motivation

Usually, to investigate a biological process, the original system is simplified by separating and isolating the involved components. The environmental conditions are optimized and newly defined in a test tube. Sophisticated methods enable to study the process of interest and often require additional modifications. Thereby, the nature of a system becomes blurred. Finally, the biggest challenge is to transfer the generated knowledge back into the context of the natural environment. This thesis attempts to develop a strategy of how a system can be studied in its natural environment. Thereby starting with an engineered molecule and going stepwise towards its natural appearance.

The genome of every living organism consists of coding regions (exons) disrupted by non-coding regions (introns); different among all is the proportion only. During protein biosynthesis, the coding regions, which contain the genetic information, are further processed and the non-coding intronic regions are excised. This process called splicing is the focus of the present thesis. We investigated a specific group IIB intron named *Sc*.ai5γ, which is located in the mitochondrial genome of baker's yeast. Its function is of particular interest because it is evolutionarily related to the spliceosomal intron present in eukaryotes. Although *Sc*.ai5γ belongs to the best-understood RNAs along all introns, its detailed function remains elusive.

This work includes the following main objectives (FIGURE 1.1)

a) Application of a new labeling strategy for large RNA

b) Development of a new exon-carrying intron derivative (eD135e-L14) and adjustment of an encapsulation procedure to this RNA system for single-molecule Förster resonance energy transfer (FRET) analysis

c) Investigations of the influence of the flanking exons on the first step of splicing in case of eD135e-L14

d) Investigations of the influence of the flanking exons on the whole splicing process in the case of *Sc*.ai5γ

e) Developing a strategy to track the splicing process in the cellular environment

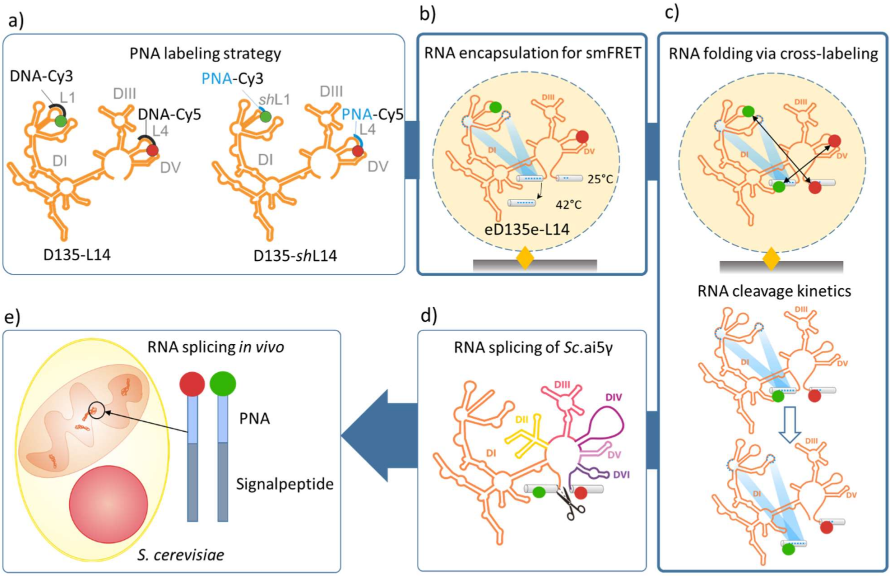


*Figure 1.1 Project overview. **a)** The starting point is the intron derivative D135-L14 and a novel labeling strategy by using PNAs (Schmitz et al., 2015). **b)** The biological relevant exons were introduced yielding eD135e-L14. The folding of eD135e-L14 was investigated applying single-molecule encapsulation. **c)** The influence of the exons on the folding pathway was further studied by cross-labeling eD135e-L14. Subsequently, the kinetics of the 5'exon cleavage were under investigations. **d)** Elucidating the whole splicing process by studying the wild-type Sc.ai5γ was the next step. **e)** The final goal is to visualize the splicing using PNAs fused to a signal peptide inside the mitochondrial matrix. **b)** and **c)** are highlighted because they are of main focus in the present thesis.*

The starting point of the present project was the engineered construct D135-L14, which is a derivative of *Sc*.ai5γ and enables to investigate the first step of splicing, folding and subsequent 5'exon cleavage. We applied a beneficial labeling strategy by using synthetic produced peptide nucleic acids (PNAs) to visualize the folding process via fluorescence (FIGURE 1.1A). The results are explained in detail in Schmitz *et al.*, 2015. PNAs turned out to be very promising labels for RNA. Subsequently, a new construct was designed. eD135e-L14 contains the biological relevant exons and is focus of the present thesis in CHAPTER II (FIGURE 1.1B). The folding of eD135e-L14 was visualized by combining single-molecule spectroscopy with RNA vesicle encapsulation. After establishing the labels and the encapsulation procedure, the folding of eD135e-L14 was studied in more detail by cross-labeling the construct (FIGURE 1.1C). In parallel, the kinetics of the 5'exon cleavage were investigated. Both, cross-labeling and cleavage kinetics of eD135e-L14 are part of CHAPTER III. Towards the natural system, the flanking exons were introduced to the wild-type intron *Sc*.ai5γ and subsequent both steps of splicing, intron excision and exon ligation were investigate (data not shown in the thesis) (FIGURE 1.1D) (Zelger-Paulus, 2013; Vukadin, 2016). Finally, a strategy of visualizing the intron respective the splicing process *in vivo* was developed by using PNAs fused to a signal peptide (FIGURE 1.1E) (Zelger-Paulus, 2013). The goal to follow splicing once in its place of origin, namely the mitochondrial matrix is still driving the motivation for this project. We hope that the knowledge gained through the biochemical and biophysical methods will help to achieve this objective.

1.2 Research background

1.2.1 General introduction to RNA

During protein biosynthesis, double-stranded deoxyribonucleic acid (DNA) is transcribed intro single-stranded messenger ribonucleic acid (mRNA), which is further translated into an amino acid chain (protein). For a long time, it was believed that the mRNA is just transferring the genetic information from the DNA towards the corresponding protein. Conspicuously was the deviation between the huge size of the genome and the relatively small number of genes responsible for coding proteins (Cech, 1971; Gregory, 2001). After the discovery of non-coding DNAs, it became clear that a significant proportion of which appears to have no biological function has been called junk DNA (Gilbert, 1978). Since then, research tries to answer the question which biological role underlies in non-coding sequences. Today, it is known that less than 1.5 % of the human genome consists of protein-coding regions. The rest consists of non-coding regions, pseudogenes (non-functional gene copies), and transposable elements (Lander *et al.*, 2001; Gregory, 2001).

Non-coding RNAs were found to be involved in many cellular processes, including RNA-splicing (small nuclear RNA), RNA-translation as part of the ribosome (rRNA) as well as for the transfer of amino acids (tRNA), and regulation of transcription and translation as riboswitches. The discovery of catalytic active RNAs was honored in 1989 with the Nobel Prize in Chemistry for T. Cech and S. Altman (Kruger *et al.*, 1982; Guerrier-Takada *et al.*, 1983). Nowadays, catalytically active RNAs are called ribozymes, enzymes entirely made of RNA (Pyle, 2016).

1.2.2 History and mystery of introns

In 1977 D.S. Hogness and co-workers discovered intragenic sequences within genes which interrupt expressed regions (Lifton *et al.*, 1977). Later, W. Gilbert named them introns while the expressed regions were named exons (Gilbert, 1978). During protein biosynthesis, the introns localized in the heterogeneous nuclear RNA (hnRNA) are removed from the maturating RNA by a process called splicing. The outcome of this RNA maturation step is messenger RNA (mRNA), which only is composed of exons (FIGURE 1.2). The mRNA is then finally translated into the corresponding amino acid sequence of the encoded protein. Although introns are widely spread among the genome, the function and biological significance of these non-coding sections remain unclear. Interestingly, the higher developed an organism is evolutionary, the more introns are found in its genome (Zelger-Paulus, 2013). The content of introns in the human genome is about 25 %, whereas it is only 4 % in baker's yeast (Jo and Choi, 2015). On average 40 % of a gene are intronic sequences (Jo and Choi, 2015). It is very puzzling why cells are putting so much effort and expensive energy in duplicating the whole gene, including the introns, which then need to be excised with the help of large RNAs and co-factors (Jo and Choi, 2015). In the review of Bong-Soek and co-authors potential functions of introns are summarized, e.g. the regulation of alternative splicing, the modulating of gene expression or the transport of mRNA (Jo and Choi, 2015). Introns are therefore definitely no junk and their importance will be revealed as insights in this rather young research field is growing.

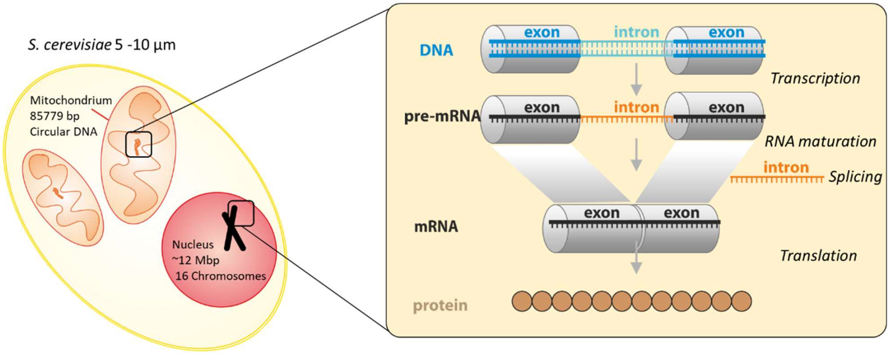


Figure 1.2 Scheme of Saccharomyces cerevisiae cell showing a mitochondrium and the nucleus (left) where the transcription and RNA maturation (right) is taking place. The double-stranded DNA containing exons and introns is transcribed into single-stranded RNA. During RNA maturation, the intronic sequences are excised by RNA-splicing. The remaining exons are ligated and translated by the ribosome machinery into an amino acid chain.

According to a recently published review of Zimmerly and co-authors, introns can be divided into four types (McNeil *et al.*, 2016). The nuclear pre-mRNA introns were found in the nucleus of eukaryotes where the splicing is supported by the spliceosome, a complex of small nuclear RNA (snRNA) and proteins (McNeil *et al.*, 2016). In contrast, the group I and group II introns are ribozymes, which undergo autocatalytic splicing. Both types are distributed among bacteria, lower eukaryotes (e.g. yeast), fungi, and plants (Saldanha *et al.*, 1993). Bulge-helix-bulge introns are short introns present in tRNA precursors, which need certain splicing factors to perform the cleavage reaction (McNeil *et al.*, 2016). The splicing reactions are similar among all four types of introns but only spliceosomal, and group II introns have major parallels (PARAGRAPH 1.2.3), which is an indication of evolutionary relation (McNeil *et al.*, 2016; Zimmerly and Semper, 2015; Fica *et al.*, 2013).

1.2.3 Group II introns

1.2.3.1 Evolution and relationship to the spliceosome

Group II introns show a huge structural and functional diversity. Through evolution, they derivatized into multiple forms adapted to individual circumstances. Although group II introns are part of genomes of diverse domains of life such as bacteria, archaebacteria, mitochondria, and chloroplasts they cannot be found in the nucleus (Zimmerly and Semper, 2015). Previous studies, however, provide evidence that group II introns are the ancestor of nuclear spliceosomal introns found in eukaryotes (Cech, 1986; Zimmerly and Semper, 2015; Fica *et al.*, 2013). Thus, understanding the underlying function and structure of group II introns is essential to elucidate the evolutionary relationship between both, and to shed more light on the spliceosomal machinery.

Structural, as well as mechanical similarities between the group II introns and the spliceosomal introns, were found. For example, recent studies from Fica and co-authors identified identical coordination of metal ions inside the active site of the spliceosome to the group IIC intron from *Oceanobacillus iheyensis* as it was known from its crystal structure (Fica *et al.*, 2013; Toor *et al.*, 2008). Even the co-factors involved in the splicing process share functional domains. The spliceosomal protein Prp8 contains a reverse transcriptase domain similar to the intron-encoded proteins (IEPs), which are involved in splicing (Galej *et al.*, 2013; Dlakić and Mushegian, 2011). All mentioned similarities and the common branching mechanism of splicing between the eukaryotic spliceosome and the group II introns support the hypothesize of common ancestors.

Classification of group II introns

Group II introns can be divided into subclasses by two different classification systems; either according to their mode of exon recognition or by the intron-encoded protein (IEP) phylogeny (Lambowitz and Zimmerly, 2011; Zimmerly and Semper, 2015; McNeil *et al.*, 2016). Initially, group II introns were assigned to IIA, IIB, and IIC introns depending on the different exon binding interactions (FIGURE 1.3). For group IIA introns, the 5'exon is recognized through two intron-binding sites (IBS1, IBS2) and the 3'exon by a third δ-δ' tertiary interaction (FIGURE 1.3). Similar is the exon binding scheme for group IIB introns, where instead of the δ-δ' interaction the 3'exon base-pairs with IBS3 (Qin and Pyle, 1999; Michels, JR. and Pyle, 1995). Group IIC introns recognize their exons via IBS1 and IBS3 (FIGURE 1.3) (Toor *et al.*, 2001). IIA and IIB introns are found in bacteria, mitochondria or chloroplasts, whereas group IIC introns are exclusively present in bacteria. More recently, two new subclasses have been identified, IIE and IIF introns (Nagy *et al.*, 2013; Toro *et al.*, 2002; Dai *et al.*, 2003; Simon *et al.*, 2008). Both types are very similar to group IIB introns, but host unique features, which makes them distinguished subclasses.

Alternatively, group II introns are classified according to the IEP phylogeny into A, B, C, D, E, F, mitochondria-like (ML), and chloroplast-like (CL). Apparently, there is a correlation between both classification systems, because introns and their IEPs have co-evolved, e.g. the bacterial group IIC intron class corresponds to the IEP-based class C (Toor *et al.*, 2001; Zimmerly and Semper, 2015). Nevertheless, IEP-lacking group II introns are not included in the latter classification system.

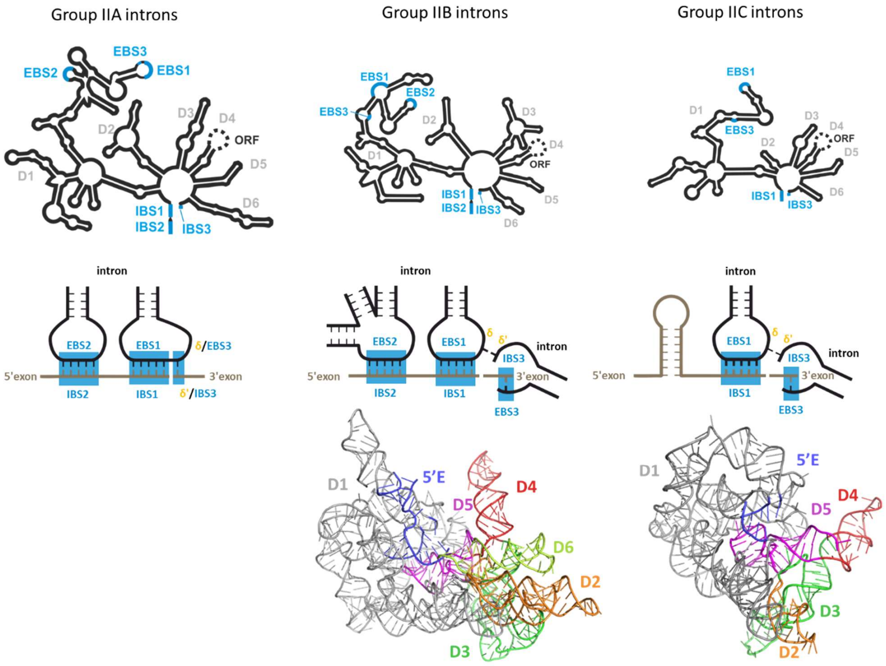


Figure 1.3 Comparison between group IIA, IIB, and IIC introns in regard to their secondary structure (top), intron-exon binding mode (middle) and tertiary structure (bottom). The intron-binding sites (IBS) and exon-binding sites (EBS) are highlighted in blue. The crystal structures of the group IIB intron from P. littoralis (PDBID 4R0D), and of the group IIC intron from O. iheyensis (PDBID 4FAQ) were rendered in Pymol. The EBSs and the 5'exon are highlighted in blue. The figure was adapted from (Zhao and Pyle, 2017) and (Pyle, 2016). The 3D figures were rendered in Pymol.

1.2.3.2 Structure of group II introns

Secondary structure

All group II introns have diverse primary structures, but share highly conserved secondary and tertiary structures. The common secondary structure of the group II introns displays six distinct domains radiating from a central wheel (FIGURE 1.4). From 5'-to 3' direction, domain D1 is the largest domain which contains the essential exon-binding sites and further recognition sites for long-range tertiary interactions (α-α', β-β', ε-ε', ζ-ζ', θ-θ', λ-λ', κ-κ') important for folding (Fedorova and Zingler, 2007; Koch, 1992; Michels, JR. and Pyle, 1995; Donghi *et al.*, 2013; Toor *et al.*, 2008). Domain D2 makes contacts with D1 and D6 and is together with domain D3 considered to be not essential for catalysis. Domain D3 is known to position D5 correctly and thereby enhances the cleavage reaction. The joining region between D2 and D3 (J2/3) is involved in placing the 3'-splice site within the active site (Podar *et al.*, 1998; Fedorova *et al.*, 2003; Robart *et al.*, 2014). Domain D4 contains an open reading frame (ORF) of about 1.5 kb encoding for an IEP. Additionally, D4 includes high-affinity binding sites for its encoded protein, but D4 is not actively involved in the splicing process. For *in vitro* experiments, D4 is often truncated to a short tetraloop, a modification which enhances the splicing reaction (Pyle, 2016; McNeil *et al.*, 2016). Domain D5 holds the most highly conserved nucleotide of the intron located at the bulge AC-motif between the two helices, which is "analogous to U6 snRNA and thereby conserved from bacteria to man" (Pyle, 2016) (FIGURE 1.4). The phosphate backbone of the two nucleotide AC bulge helps to coordinate two metal ions, which are essential for the active site formation (Sigel *et al.*, 2004; Toor *et al.*, 2008; Chanfreau and Jacquier, 1994). D5 also forms tertiary contacts to D1 (λ', κ', ζ') and folds onto the prefolded cup-form of D1 (Swisher *et al.*, 2001; Toor *et al.*, 2008). Domain D6 comprises the branch point adenosine, which initiates the nucleophilic attack of the 5'-splice site during the first step of the branched splicing pathway (explained in detail here PARAGRAPH 1.2.4) (Pyle, 2010; McNeil *et al.*, 2016). In addition, it contains tertiary interactions with D2 (π', η') (FIGURE 1.4).(Erat *et al.*, 2007; Robart *et al.*, 2014).

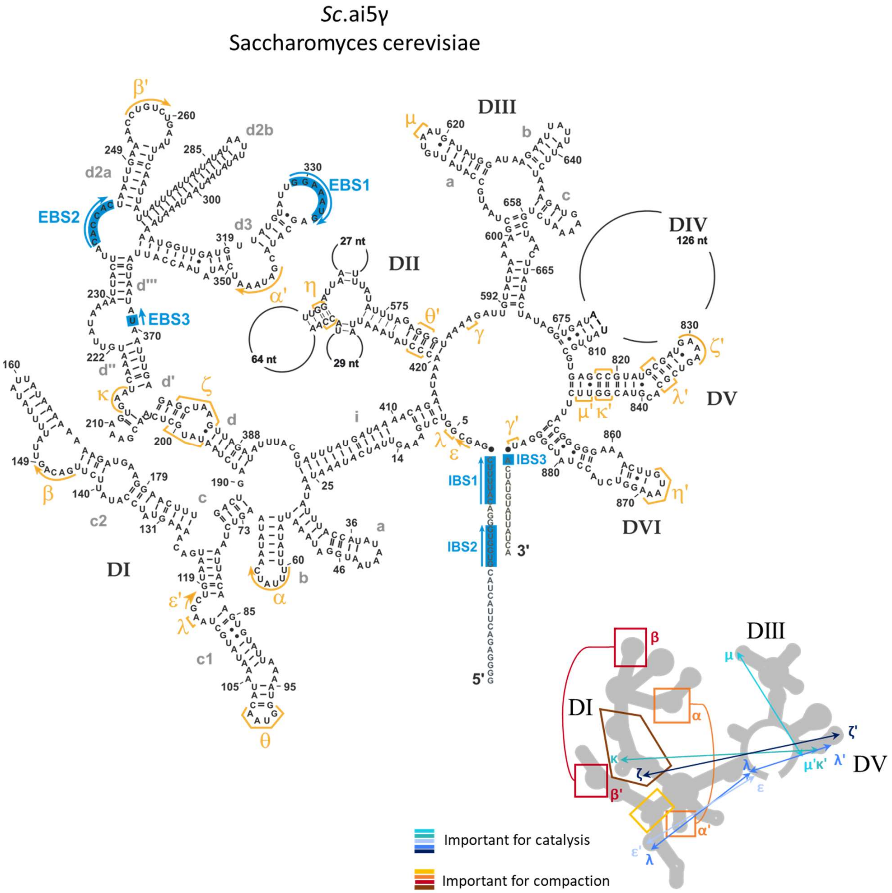


Figure 1.4 Secondary structure of the group IIB intron Sc.ai5γ. The six domains D1-D6 radiate from a central wheel-like structure. The subdivision of the individual domains is labeled in grey letters. The tertiary contacts between the different domains are highlighted in yellow and labeled by greek letters. The contacts between the intron and the exon (EBS/IBS) are colored in blue. The tertiary interactions important for catalysis are indicated in different shades of blue, whereas the regions essential for RNA compaction are highlighted in different shades of orange. Figure adapted from Fedorova et al., 2007.

Tertiary structure

After transcription, the intron RNA folds into a complex tertiary structure. Only if the RNA is folded correctly, splicing can take place. Defined conformational stages are required for the precise excision of the surrounding RNA. A milestone in structural biology and for the group II intron community was the first crystal structure of a bacterial group IIC intron from *O. iheyensis* (FIGURE 1.3) (Toor *et al.*, 2008). The 380 nt long intron comprises 80 % of the wild-type sequence and 20 % of modifications nesseccary to obtain crystallization conditions. The structure shows the interaction between the intron and its short 5'exons as well as between D1 and D5, and thereby provides important information about the hydrolytic splicing pathway (explained in detail in PARAGRAPH 1.2.4). A few years later in 2014, Robart and co-authors published the first crystal structure of the group IIB intron *P.li.*LSUI2 from the brown alga *Pylaiella littoralis* (FIGURE 1.3) (Robart *et al.*, 2014). The structure depicts an intron of 622 nt in a postcatalytic lariat stage and is until today the largest free RNA structure on the Protein Data Bank (PDB 4R0D, FIGURE 1.3). A crystal structure of a group IIA introns is not available yet.

1.2.4 Group II intron mobility

1.2.4.1 Branching pathway

Group II introns are ribozymes, which fold into a complex three-dimensional structure thereby initiating their auto-catalytic splicing during RNA maturation. The essential nucleotide in the splicing process is the branching adenosine residing in domain D6 (FIGURE 1.4). The two-step branching pathway was named after this highly conserved nucleotide and is chemically identical to the splicing process of nuclear pre-mRNA (FIGURE 1.5). *In vivo* as *in vitro,* the branching pathway follows two consecutive transesterification reactions (FIGURE 1.6). In the correctly folded RNA, which implements a proper intron-exon (EBS/IBS) interaction (FIGURE 1.3), the 5'-splice site is positioned in the active site (Peebles, 1986; van der Veen, R. *et al.*, 1986; Costa *et al.*, 2016). During the first step of splicing, the branching adenosine of D6 acts as a nucleophile; the 2'OH-group attacks the phosphodiester linkage at the 5'-splice site yielding in a lariat intron structure and a cleaved 5'exon (FIGURE 1.5) (Peebles, 1986; van der Veen, R. *et al.*, 1986). Induced by a conformational change, the 5'exon moves out of the active-site giving room for the 3'-splice site, which now moves into the catalytic center (Robart *et al.*, 2014; McNeil *et al.*, 2016). In the second step of splicing, the 3'OH-group of the 5'exon acts as a nucleophile attacking the 3'-splice site. The resulting splicing products are a lariat intron and a ligated exon. Both steps are energetically neutral and reversible, which is an important property for retrohoming (PARAGRAPH 1.2.4) and makes the group II introns mobile genetic elements (Zimmerly and Semper, 2015). Although structural changes occur in the active site during both splicing steps they are catalyzed by the same functional groups (Marcia and Pyle, 2012; Chanfreau and Jacquier, 1994; Pyle, 2016; Costa *et al.*, 2016).

In vitro, this splicing process can be induced by high Mg^{2+} concentrations and elevated temperature. Instead, *in vivo* the RNA functions as a stable ribonucleoprotein (RNP), it folds and undergoes self-excision only as an RNA-protein complex. The protein is encoded on the host intron (IEP), more precisely in the ORF of domain D4. It is a multidomain maturase protein, which binds as a dimer with a very high affinity (K_d in picomolar range) to the unspliced intron either to domain D4 or less frequently to D1 (Saldanha *et al.*, 1993; Qu *et al.*, 2016). The IEP consist of four domains, a maturase, a reverse transcriptase (RT), an endonuclease domain (EN), and a DNA-binding domain (Qu *et al.*, 2016). The maturase remains bound to the intron also after cleavage for the reverse splicing process. Introns laking an ORF and thus lacking a specific maturase are supported by other splicing factors, e.g. the ATP-dependent helicase Mss116 in yeast, whose expression underlies nuclear control (Qu *et al.*, 2016; Zhao and Pyle, 2017).

1.2.4.2 Hydrolytic pathway

The hydrolytic pathway differs from the branching pathway in that way that the nucleophilic attack of the 5'-splice site is induced by a water molecule instead of a nucleotide (FIGURE 1.5) (Jarrell *et al.*, 1988; Daniels *et al.*, 1996). The hydrolytic cleavage of the 5'exon is not reversible. Therefore introns which are cleaved by this pathway are not genetically mobile. In a second step, the exons are ligated. Also different from the branching pathway is the reaction product. The intron is excised and released in the form of a linear sequence rather than a lariat (Jarrell *et al.*, 1988; Daniels *et al.*, 1996). Zimmerly and co-authors summarized the circumstances under which the hydrolytic pathway is preferred over the branched splicing pathway: For *in vitro* self-splicing under high concentrations of salt (K^+ and Mg^{2+}) and elevated temperature (Jarrell *et al.*, 1988; Daniels *et al.*, 1996), for mitochondrial ribosomal introns, a subclass of group IIB introns with an inserted 5'-extension (Li *et al.*, 2011), and for *in vivo* splicing of organellar introns lacking D6 (Podar *et al.*, 1998; McNeil *et al.*, 2016). In addition, the hydrolytic activity of a so-called minimal ribozyme construct consisting of only two domains, D1 and D5, was shown. Here, only the first step of splicing was visualized by adding in *trans* a substrate consisting of a short 5'exon and the first 7 nucleotides of the intron (Michels, JR. and Pyle, 1995).

Both splicing pathways, the hydrolytic and the branching pathway, can occur in parallel *in vitro* and *in vivo* (Daniels *et al.*, 1996). Which splicing pathway is preferred depends on the surrounding conditions. It is known, that *in vitro*, different metal ions promote one pathway over the other, e.g. the hydrolytic pathway is favored in the presence of potassium ions while the branching pathway requires ammonium ions (Jarrell *et al.*, 1988; van der Veen, R. *et al.*, 1986; Daniels *et al.*, 1996).

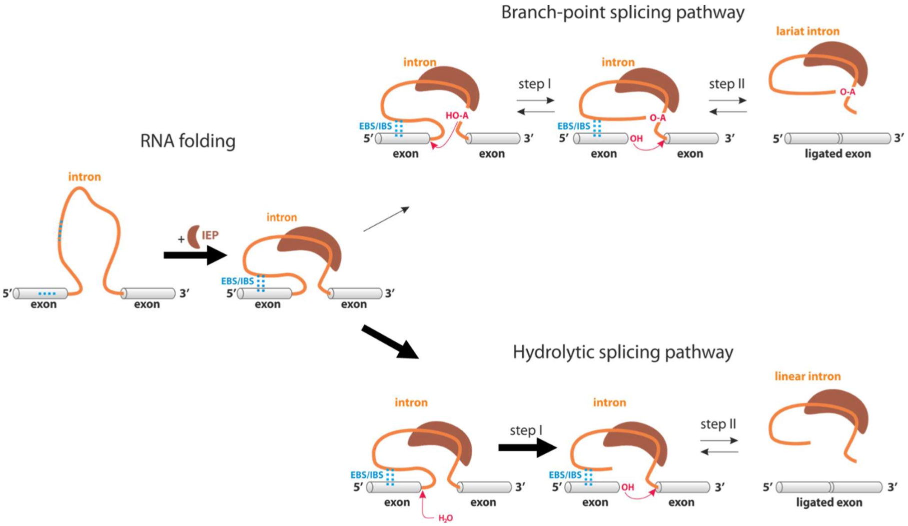


Figure 1.5 The two possible splicing pathways, hydrolytic and branch-point, of group IIB introns. In vivo, splicing is supported by an intron-encoded protein (IEP, brown) whereas in vitro splicing is induced by high salt concentration and elevated temperature. Both pathways can occur in vivo and in vitro depending on the environment. The derivatives of Sc.ai5γ used in this study, D135, D135-L14, and eD135e-l14, undergo only the first step of splicing via the hydrolytic pathway, which is the cleavage of the 5'exon (broad arrows). The figure was adapted from (Daniels et al., 1996) and (Zingler et al., 2010).

1.2.4.3 *Chemical background of splicing reaction*

The chemical reaction of the splicing process is a metal ion assisted catalysis in both, group I and group II introns (Figure 1.6). Crystal structures of the group IIC intron *O. iheyensis* at different stages of catalysis revealed the active-site formation and how metal ions are involved (Marcia and Pyle, 2012). The backbone of the AC-bulge in domain D5 creates a negatively charged binding pocket for two magnesium ions. The two divalent metal ions position both the scissile phosphate of the splice site as well as the reactive nucleophile. Additionally, two potassium ions play a crucial role in stabilizing the cleaved 5'exon during the whole splicing process (Marcia and Pyle, 2012, 2014).

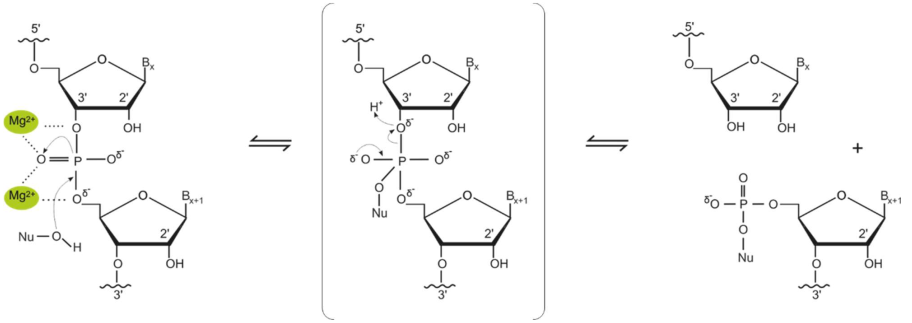


Figure 1.6 Mechanism of RNA-cleavage. The bridging phosphate between two nucleotides of the exposed splice site is attacked by a nucleophile (Nu). Depending on the intron and reaction step this nucleophile can be either a 3'-OH of a guanine (group I introns), a 2'-OH of an adenine (group II introns, branched pathway), a 5'-OH of the excised 5'exon (second step of splicing) or a water molecule (hydrolytic pathway).

1.2.4.4 *Structural stages during RNA splicing*

The group II intron is a dynamic system, which undergoes multiple structural stages during the splicing process. With the help of the above mentioned 3D crystal structures, the splicing pathway could be reconstructed by snapshots of structures showing the intron in relevant splicing stages. In her review from 2016, Pyle provides a comprehensive overview of the splicing mechanism including five structural stages: being first the precatalytic state, second the step 1 state, third the active-site rearrangement between the steps 1 and 2, fourth the step 2 state, and last the post-catalytic state (Pyle, 2016). The precatalytic state is captured crystallographically by a group IIC intron (*O. iheyensis*), which was trapped in this stage by inhibiting the splicing via calcium ions (Toor *et al.*, 2008).

A comparison of the active-site configuration between the precatalytic state and the post-catalytic state and the intron-exon complex revealed no structural differences, which indicate an identical conformation during step 1 and 2 states (Marcia and Pyle, 2012). The crystal structure of the group IIC intron from *O. iheyensis* in the presence of its 5'exon visualized the step 1 state PDBID 4FAR, (Marcia and Pyle, 2012). According to this structure, after 5'exon cleavage, the exon stays bound to the EBS1 site, and a potassium ion is directly involved in the stabilization of the step 1 product (Pyle, 2016). The active-site of the intron undergoes structural changes between step 1 and 2 states (Robart *et al.*, 2014; McNeil *et al.*, 2016). 5'exon cleavage induces the entry of the 3'-splice site into the active site and its correct positioning for the attack of the 5'exon. Rearrangement of the active site also includes displacement of domain D6 containing the branched adenosine. There is no crystal structure of the intermediate state between both splicing steps. Only a potential intermediate state was crystallized by replacing potassium ions by sodium and lithium ions and trapping the intron after the first splicing step (Marcia and Pyle, 2012). No crystallographic data of the second step of splicing are available so far. From cross-linking, chemical mapping and mutagenesis experiments it is known that the second splicing step always follows the first one and that the confirmation of the active-site is unchanged before and after splicing took place (Chanfreau and Jacquier, 1994; Pyle, 2016).

The last stage of splicing, the postcatalytic state was investigated comprehensively by crystallization the linear group IIC intron bound to its 5'exon on the one hand and of the free group IIC intron (*O. iheyensis*) on the other hand (Chan *et al.*, 2012). The comparison between both structures revealed no conformational change of D5. Further, it was shown that the catalytic metal ions, two magnesium (II) and two potassium (I) remain in the active-site. The postcatalytic stage of the branching pathway of *P. littoralis* showed a lariat intron in which D6 is far away from the active site.

1.2.4.5 Retrohoming

The majority of group II introns present in bacteria are located outside of housekeeping genes and behave as retroelements (Dai and Zimmerly, 2002). In contrast, organellar introns are always resident in housekeeping genes and act as splicing-only introns (Simon *et al.*, 2008). Those introns also lack the ORF for an IEP. The group II intron-IEP complex is responsible for the recognition of a specific DNA motif. Upon binding the DNA, the IEP unwinds the double-strand and reverse splicing of the intronic sequence takes place. The template strand is cleaved by the En domain, thereby produces a primer for the reverse transcription of the intron (McNeil *et al.*, 2016).

1.2.4.6 Obscure splicing pathways

Other splicing pathways, like the spliced-exon reopening (SER), the *trans*-splicing or the formation of intron circles occur less frequently and are less understood (FIGURE 1.7). The SER was discovered in 1987 by Jarrell and co-authors. It is an unproductive side reaction of splicing pathway discovered *in vitro* and observed under all splicing conditions (Jarrell *et al.*, 1988; Daniels *et al.*, 1996) . The ligated exons are rapidly cleaved at their junction by a lariat intron, which finally results in a cleaved 5'- and 3'exon. Some organellar introns, e.g. the intron *Ll*.LtrB from *Lactococcus lactis* can undergo *trans*-splicing, which means that the intron is encoded in several RNA regions that are subsequently fused to one (Belhocine *et al.*, 2007). The formation of intron circles *in vivo* was first reported in 2001 by Murray and co-authors (Murray *et al.*, 2001; Molina-Sánchez *et al.*, 2006). It is hypothesized that they occur from a *trans*-splicing process, but their biological role remains unclear (Murray *et al.*, 2001; Molina-Sánchez *et al.*, 2006).

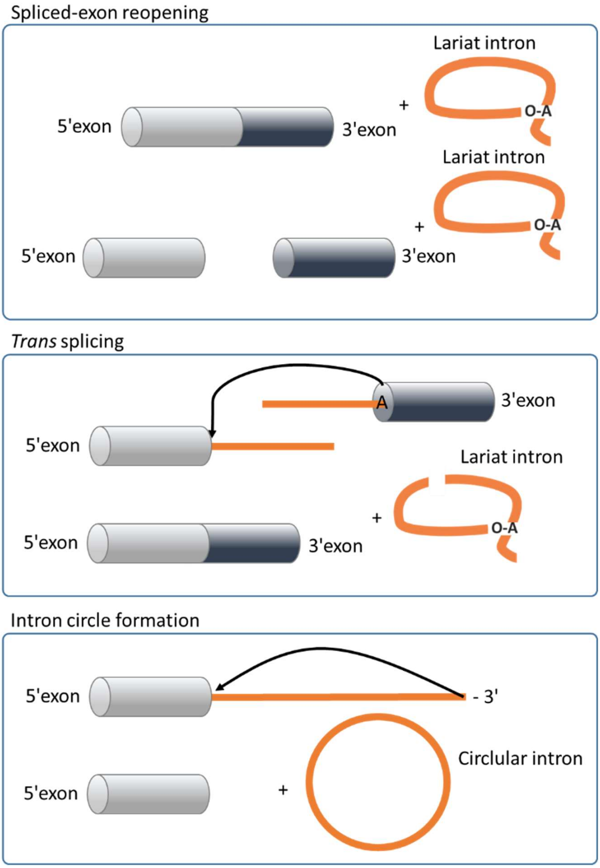


Figure 1.7 Obscure splicing pathways. In vitro occurring spliced-exon reopening. The exon junction is spliced by a lariat intron. During trans splicing , the intron is encoded on more than one RNA regions. Formation of intron circles. The 3'end attacks the 5'-splice site and induces splicing, yielding in a circle and a spliced 5'exon. Figure was adapted from (McNeil et al., 2016).

1.2.5 The group IIB intron *Sc*.ai5γ and its derivatives

The model system of the present thesis is the group IIB intron *Sc*.ai5γ located in the mitochondrial genome of *Saccharomyces cerevisiae*. More precisely, it is the fifth intron of the *cox*1 gene, which encodes for the subunit 1 of the cytochrome c oxidase. This transmembrane protein plays an important role in the respiratory chain of the mitochondrion. Since *Sc*.ai5γ is residing in a housekeeping gene, it lost its mobility and does not contain an ORF. The folding and splicing *in vivo* are supported by the DEAD-box helicase Mss116. *Sc*.ai5γ can undergo the hydrolytic and branching splicing pathway *in vitro*, depending on the type of metal ion present, but it remains unclear whether both splicing pathways exist *in vivo*.

"In order to study a single, defined conformation of the active intron, a ribozyme deletion construct was created" in 2001 by Swisher et al. (Swisher *et al.*, 2001). This construct is a derivative of *Sc*.ai5γ, composed of only three domains D1, D3, and D5 (FIGURE 1.8). This generated construct, D135, can undergo the first step of splicing which was studied in the presence of an added substrate. About 83 % of this ribozyme is active under standard conditions with a rate constant of 0.48 min^{-1} (Swisher *et al.*, 2001). Precise intron splicing requires accurate RNA folding. The simplified construct D135 was further modified to investigate the folding mechanism of *Sc*.ai5γ (Steiner, 2008). The intron derivative D135-L14 contains labeling platforms L1 and L4 in domain D1 and D4. It has been shown that the cleavage activity drops down by 50 % in the presence of these modifications and their labels (Fiorini, 2015). In this study, D135-L14 was further modified by adding the biological relevant exons at the flanking sites to follow the splicing reaction in *cis*. The resulting construct eD135e-L14 is labeled fluorescently by peptide nucleic acid (PNA) labels, which carry each a fluorophore and hybridize to the flanking exons.

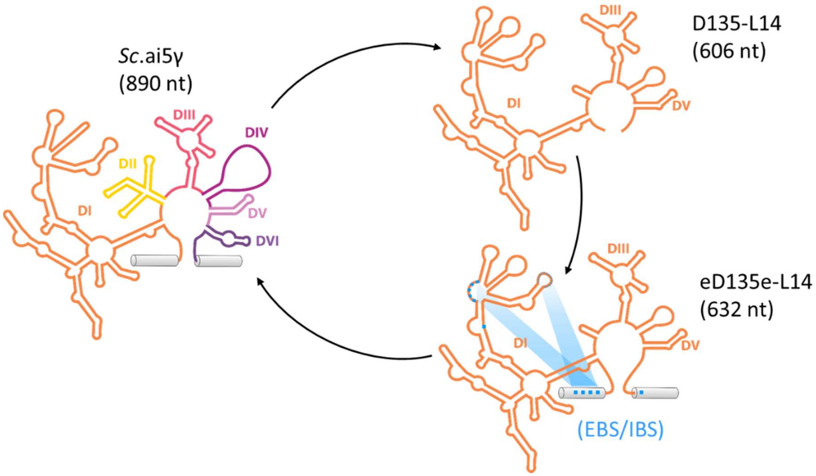


Figure 1.8 From the original wild-type group IIB intron Sc.ai5γ including 6 domains and flanking exons towards a simplified minimal active construct D135-L14, consisting of only D1, D3, and D5 and additional labeling platforms in L1 and L4. To investigate the influence of the flanking exons on the folding, two short exons were introduced resulting in the construct eD135e-L14.

1.2.5.1 Folding of *Sc.ai5γ* and its derivatives

Sc.ai5γ is a large and complex ribozyme, which exhibits a hierarchical scaffold-based folding mechanism. Domain D1 is the first domain of the intron being transcribed *in vivo*, thus folding of D1 is the first step in the intronic folding pathway (Qin and Pyle, 1997). The correct fold of D1 is the rate-limiting step and provides a scaffold to which the other domains and exons are docking (Swisher *et al.*, 2002). The catalytically important tertiary interactions κ-κ' and ζ-ζ' between D1 and D5 were found to be crucial to position D5 correctly (FIGURE 1.4) (Waldsich and Pyle, 2007). As soon as the intron is correctly folded, it acquires splicing activity. Thereby, the splicing factor Mss116 supports the formation of the native fold rather than unwinding misfolded states as it was previously assumed (Zingler *et al.*, 2010). The truncated group IIB intron D135 became an important model system to study RNA folding. For the first time, it was shown that D135 as a multi-domain construct folds homogenously and in a cooperative two-state folding manner without kinetic traps (Su *et al.*, 2003). It was shown by time-resolved hydroxyl radical footprinting that D135 folds slowly, but directly to an active conformation by which 40 % of the RNA is solvent-protected (Swisher *et al.*, 2002). Thereby, the folding of D1 constitutes an intermediate folding state towards the native tertiary structure. The overall folding pathway was first described as $U \leftrightarrow I \leftrightarrow N$, whereas U is the unfolded RNA consistent of secondary elements that form only in the presence of K^+, *I* is an intermediate state in which D1 is folded, this state formation requires additionally Mg^{2+} and *N* is attributed to a native folded state which is catalytically active (Waldsich and Pyle, 2007). The hypothesis that D135, as a large structurally complex RNA folds without any incorrectly folded intermediates smoothly into its native state is still under discussion. There is evidence for a potentially hidden intermediate, e.g. the turnover of splicing reaction is only between 60 and 80 % and not 100 % as expected for a straight two-step process (Swisher *et al.*, 2002; Su *et al.*, 2005; Su *et al.*, 2003; Swisher *et al.*, 2001; Zingler *et al.*, 2010; Zingler, 2014). Single-molecule FRET (smFRET) studies on the surface-immobilized D135-L14 revealed a folding pathway from the unfolded RNA (*U*) towards the native structure (*N*) through two folding intermediates (*I* and *F*) (FIGURE 1.9) (Steiner, 2008; Steiner *et al.*, 2009; Karunatilaka *et al.*, 2010; Fiorini *et al.*, 2015). The presence of the substrate and the splicing co-factor Mss116 further stabilizes the conformation of *N* (Karunatilaka *et al.*, 2010). Recent smFRET studies revealed even five FRET states; an unfolded *U*, two intermediates I_1, I_2, a folded state *F* and a native state *N* (Fiorini *et al.* in preparation, 2018).

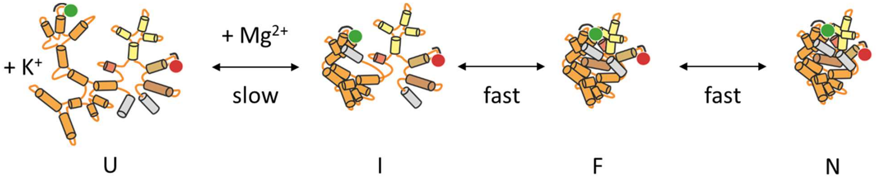


Figure 1.9 Hypothesized folding pathway of the group IIB intron D135-L14. The unfolded state (U) consists of secondary elements. The first and rate-limiting step of the RNA folding is the slow collapse of D1 (I) in the presence of Mg^{2+} attributed to an extended intermediate state. Then the intron folds fast towards its native conformation (N) through a folded intermediate (F). The folding pathway is fully reversible. The figure was adapted from (Steiner, 2008).

1.2.6 Structure meets function

1.2.6.1 *Kinetic characterization of group II intron splicing*

The catalytic activity of D135 and its derivatives were investigated by following the cleavage of a substrate by applying gel electrophoresis (Fedorova, 2012; Zingler, 2014). The substrate consists of 17 nucleotides from the 5'exon including both IBSs and the first 7 nucleotides of the intron and is given in *trans* to the pre-folded ribozyme (Michels, JR. and Pyle, 1995; Griffin, JR *et al.*, 1995). The cleavage reaction is conducted either under single-turnover conditions, meaning an excess of RNA over the substrate or the other way around under multiple-turnover conditions (Steiner, 2008;

Swisher *et al.*, 2001). In both cases, the substrate is radioactively labeled, and its cleavage is visualized using denaturing polyacrylamide gels. The kinetics of substrate cleavage are determined at different magnesium concentrations. A drawback of this method is that only the substrate is visualized and not the ribozyme or other splicing products themselves. Thus other splicing products like different ribozyme conformations cannot be observed.

Another approach to investigate the splicing activity of group II introns is to take advantage of their modularity and to assemble individual domains to functional intron derivatives. For example, a construct consisting of the 5'exon and domain D1-D3, eD123 functions as a substrate and is given in *trans* to D56, a construct made only of D5 and D6 (Erat and Sigel, 2008). In this case, the ribozyme was radioactively labeled at the 5'-end.

In the present thesis, we are applying a fluorescent labeling approach and are thus able to investigate both, the splicing activity using native PAGE and the folding pathway by applying single-molecule FRET (PARAGRAPH 1.2.6.4).

1.2.6.2 RNA labeling

Fluorescence imaging of proteins or nucleic acids is a standard technique in biochemistry and biophysics to visualize, e.g. binding events or structural rearrangements. The covalent attachment of fluorescent dyes at specific sites is easier for proteins than for nucleic acids since the twenty amino acid sequence provides more possibilities for site-specific labeling than the four nucleotides do. Nucleic acids below a size of 100 bases are synthesized chemically; thereby the required modified fluorescent moiety can be incorporated at any position. A drawback of solid-phase synthesis is that the longer a sequence is, the more error-prone it is and the lower the yields are. Even more challenging is the labeling of large nucleic acids (> 100 nt). In this case, hybridizing probes with fluorescent dyes are commonly applied to label the molecule of interest non-covalently (Armitage, 2011; Smith *et al.*, 2005) or our approach, to label an RNA of a size of 632 nt fluorescently and investigate structural changes, we applied complementary oligonucleotides bearing the appropriate dyes, which bind to the introduced extensions (Smith *et al.*, 2005). Currently, the use of short DNA-sequences of at least 18 nt is common, to ensure high labeling efficiency. The choice of the labeled position needs structural knowledge about the molecule of interest; the modification should not interfere with important functional structures. For the group II intron derivative D135, two artificial loops (L1 and L4) were introduced into protruding domains D1d2b and D4 (FIGURE 1.8)

(Steiner, 2008). These loops serve as labeling platforms for the complementary DNA (18 nt) oligonucleotides (FIGURE 1.10). Such internal labeling is very useful but at the expenses of RNA folding along with RNA activity. Indeed, a comparison of rate constants of D135 and D135-L14 shows that the ribozyme activity drops down by 50 % in the presence of the modifications and labels (Swisher *et al.*, 2001; Fiorini, 2015). Thus, smaller or non-artificial modifications are desirable. Recently, a new technology of directly labeling RNA was developed but still bears some disadvantageous as complicated chemistry and low labeling efficiency (Egloff *et al.*, 2016; Zhao *et al.*, 2017).

Alternative labeling strategies rely on hybridization probes with an unnatural backbone. Peptide nucleic acids (PNAs) are an excellent and already widely used example of being a very convenient instrument for targeting RNA *in vitro* as well as *in vivo* (Chinnery *et al.*, 1999; Pellestor *et al.*, 2008; Kummer *et al.*, 2011, 2012). PNA is a synthetic nucleic acid analog invented in the early nineties by Peter Nielsen (Nielsen, 1991). It is composed of DNA-like nucleobases, which are linked by an uncharged peptide backbone (FIGURE 1.10). The unique composition offers three main advantages over DNA: i) PNA binds RNA with a higher affinity, ii) PNA-binding specificity is higher, PNAs are more sensitive to sequence mismatches, and iii) PNAs are resistant to nuclease and protease degradation (Robertson *et al.*, 2006; Muratovska *et al.*, 2001; Egholm *et al.*, 1993). In context with RNA labeling, these benefits of PNAs enable to shorten the artificial modifications and increase the labeling efficiency.

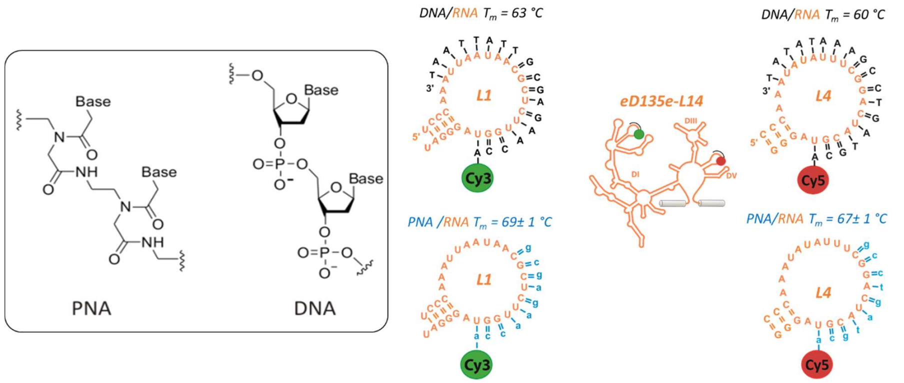


Figure 1.10 RNA labeling strategy. D135-L14 consists of two artificial labeling platforms L1 and L4 to which the complementary DNA- or PNA-strands hibridize. The melting temperature of the 18 nt long DNA oligonucleotide is lower than of the PNAs, which is 10 bases long. PNA is a nucleic acid analog composed of DNA-like nucleobases which are linked by a peptide backbone.

In the manuscript of Schmitz et al. from 2015, we applied this new PNA labeling strategy to our RNA model system to study the group II intron splicing mechanism of D135-L14 by smFRET (Schmitz *et al.*, 2015). The in-house synthesized PNAs are complementary to the introduced loops L1 and L4 (FIGURE 1.10). The sequence of the loops in D135-L14 was shortened and adjusted to the size of the PNAs. Furthermore, the PNA labeling strategy was used in the present thesis to label the flanking exons in eD135e-L14 (CHAPTER III).

1.2.6.3 Fluorescent dyes

Cyanine dyes composed of a polymethine bridge between two heterocycles are commonly used to label nucleic acids fluorescently. With increasing number of carbon atoms in the polymethine chain, the absorbance wavelength increases. The absorption maximum of Cy3, is at 548 nm while for Cy5 at 646 nm (FIGURE 1.11, TABLE 1.1). Additionally, sulfo groups at the aromatic rings increase the water solubility (FIGURE 1.11) (Mujumdar *et al.*, 1993). Cyanine dyes are in particular suitable to label nucleic acids because they are not quenched by nucleobases (Torimura *et al.*, 2001). The sulfonated cyanine dyes Cy3 and Cy5 are used in the present study as a FRET pair to investigate RNA dynamics. It is important to notice that both dyes can undergo *cis-trans*-isomerization (Sanborn *et al.*, 2007; Harvey and Levitus, 2009). Thereby, the *cis*-state is not fluorescent. The manner of interaction between the fluorescent dyes and the RNA depends on the environment. Dye staking to the RNA in either conformation, *cis*- or *trans*, significantly influences the FRET efficiency (Steffen *et al.*, 2016).

Alternative cyanine dyes like Cy3B with a rigid backbone that prevents isomerization should be considered for future smFRET measurements (Sanborn *et al.*, 2007).

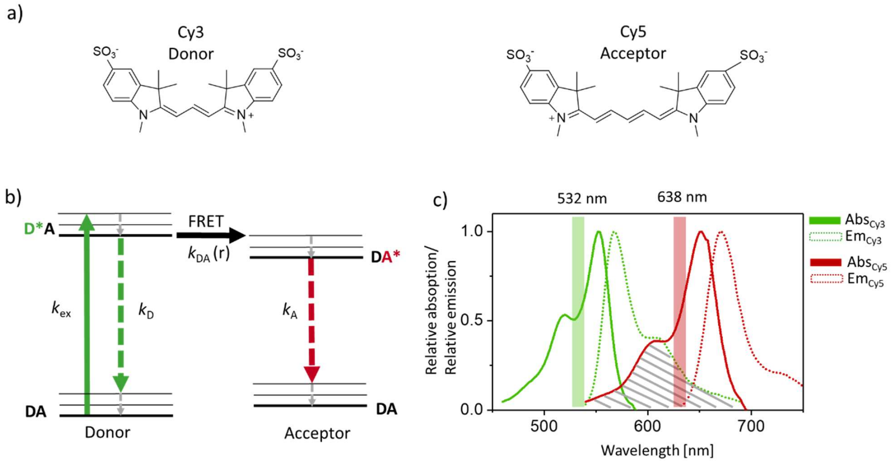


Figure 1.11 Cyanine dyes Cy3 and Cy5 and the principal of FRET. ***a)*** *The FRET pair Cy3 and Cy5 were used in this study for labeling RNA. The Förster radius $r_{Cy3,Cy5}$ = 54 Å.* ***b)*** *The simplified Jablonski diagram depicts the underlying principles of FRET between a donor (D, Cy3) and an acceptor (A, Cy5). The donor (D) initially present in the ground state is excited by a light source (green arrow) into the excited state (D*). In the excited state, D is either transferring the energy to A* in a distance-dependent manner (black arrow) or emits a photon by fluorescence (green dash arrow). The excited acceptor A* emits than a photon by fluorescence (red dash arrow). The grey arrows display relaxation from a higher singlet state to the ground state. The transition rate k_{DA} is distance-dependent (r) and determines the FRET efficiency.* ***c)*** *The absorption (Abs) and emission (Em) spectra of Cy3 and Cy5 are shown and the wavelength at which they were excited by a laser during smFRET studies. Their spectral overlay is indicated in grey. The figure was adapted from (Ha, 2007).*

The fluorescence quantum yield ϕ_F and lifetime τ_F are the most important properties of a fluorophore. The quantum yield is defined as the number of photons emitted relative to the number of photons absorbed. The higher the quantum yield, the brighter the fluorophore. The fluorescence lifetime describes the time a fluorophore spends on average in the excited state before returning to the ground state. Both, fluorescence quantum yield and lifetime are strongly dependent on the surrounding of the dye. Whether a dye is covalently bound to a residue or freely rotates influences significantly its photophysical properties, as well as the temperature does (Steffen *et al.*, 2016; Klehs *et al.*, 2014). The FRET pair Cy3/Cy5 bear excellent characteristics for smFRET, both dyes are photostable in an oxygen-free environment, both quantum yields are similar, the spectral separation between both is large, and they are commercially available (Ha, 2007).

Table 1.1 Photophysical properties of the cyanine dyes Cy3, Cy5, and Cy7 in water. Data were taken from Lumniprobe GmbH.

Fluorescent dye	Excitation maximum λ [nm]	Emission maximum λ [nm]	Fluorecsence quantum yield ϕ_F	Fluorecsence lifetime τ_F
Sulfo-Cy3	548	563	0.10	0.25[a]
Sulfo-Cy5	646	662	0.28	1.0[a]
Sulfo-Cy7	750	773	0.30	0.6[a]

[a)] (Klehs *et al.*, 2014)

1.2.6.4 Förster resonance energy transfer

Förster resonance energy transfer (FRET) is a non-radiative transfer of energy between two nearby fluorophores, a donor and an acceptor. It is widely used as a spectroscopic technique to measure distances (Schuler et al., 2005). Thereby, a donor molecule in the excited state transfers its energy to an acceptor fluorophore through the interaction of two induced dipoles. This process is depending on the distance between both dyes, the overlap of their excitation and emission spectra as well as on the orientation of the two dipole moments. The FRET efficiency E is described by the following equation

$$E = \frac{1}{1+\left(\frac{R}{R_0}\right)^6} \quad \wedge \qquad \text{Equation 1.1}$$

where R is the inter-dye distance and R_0 is the so-called Förster radius; the distance between the donor and the acceptor at which the FRET efficiency is 50 %, R_0 = 54 Å for Cy3 and Cy5. FRET results in a decrease of fluorescent intensity of the donor and simultaneously in an increase of acceptor fluorescence.

Single-molecule FRET was first applied by T. Ha in 1996 and has evolved since that to a powerful microscopic technique to investigate dynamic systems (Ha *et al.*, 1996). The advantages of using single-molecule over ensemble FRET are elucidating a heterogeneous system, no need for molecule synchronization, and the detection of rare occurring intermediate states (Ha, 2001). There are two different ways of detecting single molecules, by confocal or wide-field microscopy. Thereby, the molecule of interest can be either anchored on a surface of diffuse freely in solution. For the present smFRET study, we are applying wide-field microscopy on surface-immobilized RNA molecules using total internal reflection (TIR) excitation. Thereby, the incident laser is totally reflected at the interface between the quartz surface and the specimen and generates an electromagnetic or evanescent field of about 200 nm (FIGURE 1.12) (Stanley, 2011; Fish, 2009). Because the intensity of the evanescent light decays exponentially with distance, only fluorophores close to the surface get excited. The advantage of such a small excitation window is a reduced fluorescence background which increases the signal-to-noise ratio (Fish, 2009). The emitted photons from the donor and the acceptor are detected individually on an electron-multiplying charged-coupled device (EMCCD). The fluorescence signal needs to be corrected for crosstalk between the two detection channels; the direct excitation of the acceptor by the donor laser and the bleed-through of the donor signal into acceptor channel. After correction of the background and crosstalk, the apparent FRET efficiency is calculated accordingly

$$FRET = \frac{I_D}{I_D + I_A} \qquad \text{Equation 1.2}$$

where I_D and I_A are the donor and acceptor intensities. The apparent FRET provides only approximate distances, because further corrections for e.g., the orientation of the dyes κ^2, the different quantum yields, and the detection efficiency are required (McCann *et al.*, 2010).

smFRET experiments on the folding of D135-L14 were carried out on surface-immobilized molecules using TIRFM (FIGURE 1.12). A sequence elongation on the 3'end was used to hybridize a biotinylated oligonucleotide. The surface was coated with biotinylated-PEG (polyethylene glycol) and subsequent streptavidin to which the biotinylated RNA binds. Such direct molecule immobilization can lead to non-specific interactions of the molecule with the surface (Boukobza *et al.*, 2001). Another important fact is that tethering a molecule on a rigid position influence its flexibility and thus can have an effect on the folding. A new immobilization technique that overcomes molecule-surface interactions and the problem of sterical hindrance through molecule anchoring was developed by which single molecules are trapped inside tethered vesicles (Boukobza *et al.*, 2001). The vesicles themselves are partially biotinylated and thus immobilized onto a surface while the encapsulated molecule can freely diffuse inside (FIGURE 1.12). This promising encapsulation technique is already in use for single-molecule studies of, e.g. protein-DNA, protein-protein interactions, or binding events of short mismatched DNA sequences (Cisse *et al.*, 2012; Lamichhane *et al.*, 2010).

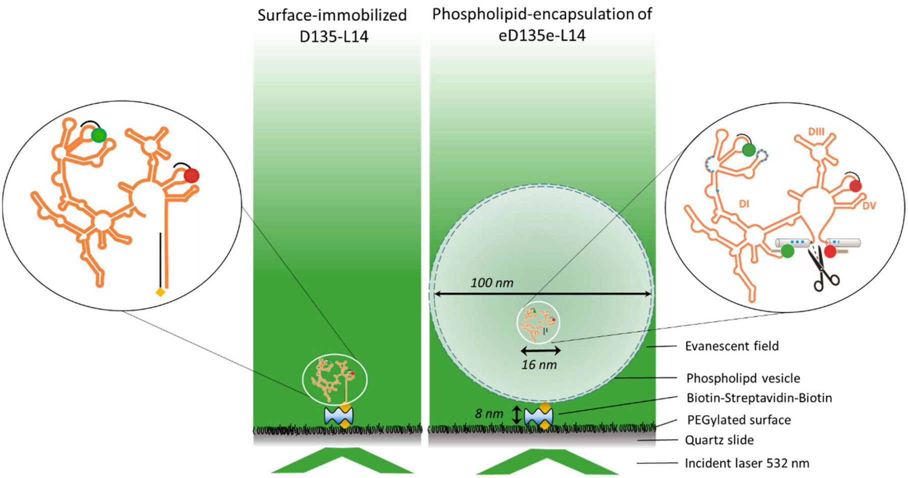


Figure 1.12 Both immobilization techniques in comparison. Classical surface immobilization of D135-L14 (left) versus vesicle encapsulation of eD135e-L14 (right). D135-L14 is directly tethered via its biotinylated 3'-oligonucleotide on a streptavidin-biotin-PEG surface (left) whereas eD135e-L14 is freely diffusing in phospholipid vesicles (right). 5'exon cleavage can be induced by elevated temperature. eD135e-L14 can be labeled fluorescently in the loops L1, L4 and at both exons.

In the present study, we are encapsulating the large ribozyme eD135e-L14 and follow its folding by smFRET (TIRF-based) induced by temperature (Figure 1.12). First, we show the influence of encapsulation on the RNA dynamics and proof that encapsulation is a suitable method for our model system (Chapter II). The classical direct immobilization of eD135e-L14 via the exons is not possible because both exons are involved in folding and are excised, which would lead to loss of surface contact. Thus, the folding of the exon-carrying intron (eD135e-L14) can only investigate by trapping it in vesicles (Chapter II). Further, we label eD135e-L14 at different positions to get more insights into the individual domains and how they are involved in the folding process (Chapter III). We then correlate the apparent FRET to approximate distances with the help of the *de novo* modeled structure of eD1345.

2. Chapter

Following RNA folding by smFRET through encapsulation

Susann Zelger-Paulus, Mélodie C.A.S. Hadzic, Richard Börner, Besim Fazliji, Roland K.O. Sigel

2.1 Introduction

Single-molecule (SM) spectroscopy stands out for revealing cross-sample variability, the distribution of subspecies, heterogeneity of structures and dynamics as well as the separation of reaction intermediates or subpopulations usually averaged out in ensemble measurements (Kowerko *et al.*, 2015; Börner *et al.*, 2016).

Förster resonance energy transfer (FRET) is a non-radiative transfer of energy between two nearby fluorophores, a donor and an acceptor. It is widely used as a spectroscopic technique to investigate binding events or structural rearrangements. Single-molecule spectroscopy in combination with FRET is a powerful technique to investigate, e.g. dynamic processes like a folding pathway, which obscures fast transitions or rare intermediate folds. smFRET measurements can be carried out by applying two different detection methods, confocal or widefield microscopy.

The classical confocal smFRET experiments are carried out on free diffusing molecules through a defined detection volume (~1 fL). Therein, single-molecule sensitivity is achieved by combining a very small detection volume and a low concentration of the molecules under study. Thereby, the probability of two molecules in focus is low. The observation over a long time range yields an average of many molecules. Thus, this method is a quasi-ensemble measurement. This point-detection method enables a very high time resolution, which allows detecting dynamics far below 1 ms (Schuler and Hofmann, 2013). Another beneficial aspect is that fluorescence anisotropy and lifetime can be measured at the same time (Time-correlated single photon counting acquisition, TPSPC) and give insights into the photophysics of the fluorophores (Schuler *et al.*, 2005; Schuler and Eaton, 2008; Roy *et al.*, 2008; Lee *et al.*, 2010). Interconversion kinetics between two distinct FRET states that take place on a longer time range than the observation time, which is the time a molecule diffuses through the focus volume (~1 ms), cannot be followed by this method (Schuler and Hofmann, 2013).

In contrast to confocal smFRET, applying a widefield microscope requires the immobilization of the molecule of interest. An advantage of immobilizing molecules onto a surface is the parallel observation of many molecules over a long period of time via camera detection. The observation time is limited by the photobleaching of the dyes. This method allows to follow conformational transitions of one single-molecule over time as long as kinetics are within the time resolution of the camera and the observation time. The limitation of time resolution by the laser illumination and camera detection can be overcome by applying a stroboscopic excitation. On the other hand, molecules need to be immobilized onto a surface, which can influence the folding and the activity of the molecule under study. For example, the position at which a molecule is tethered, e.g. termini or a protruding domain can be essential for the activity. In addition, interactions of the molecule or the fluorophores with the surface, which can lead to a misfolding or prevent the formation of the native structure is a decisive disadvantage of this method. Also, the time resolution is limited by the readout of the camera, which is about 1 ms.

An elegant way of circumventing the disadvantages of direct immobilization and simultaneously investigating dynamics over prolonged observation periods is to use surface-attached vesicles which carry the molecule of interest (Cisse *et al.*, 2007). The vesicles are usually very small (~100 nm) and anchored on the surface (FIGURE 2.13). Thus, the exponential decay of the evanescent field (~200 nm) within the vesicle diameter is negligible, and the excitation probability can be regarded as constant within the vesicle (Zelger-Paulus, in preparation, 2018). The molecules freely diffuse inside the vesicle, and only the exchange of ions through the pores is possible (Cisse *et al.*, 2007). Further, encapsulating catalytically active biomolecules allows to follow not only folding processes but also catalytic reactions as both, reactants and products, are trapped inside the vesicle.

The group IIB intron *Sc*.ai5γ is a self-cleaving ribozyme located within a housekeeping gene of the mitochondrial genome of *S. cerevisiae*. The excision of the non-coding intron from the surrounding is called splicing and results in a mature mRNA. During the splicing process the intron RNA folds into a defined 3D-structure, which is catalytically active and leads to self-cleavage and a subsequent ligation of its flanking exons. Splicing *in vivo* under physiological salt concentrations (i.e. 8 mM $MgCl_2$) requires the support of chaperone proteins, e.g. the ATP-dependent helicase Mss116 (Halls *et al.*, 2007; Zingler *et al.*, 2010). In contrast, *in vitro* and in the absence of any protein, self-splicing can be induced by high salt concentrations (500 mM KCl, > 40 mM $MgCl_2$) and elevated temperature (42 °C). Splicing is a multistep process, to be able to study individual steps, *Sc*.ai5γ has been simplified (Swisher *et al.*, 2002; Qin and Pyle, 1999). The resulting derivative D135 contains domain D1 and D5; both are essential for catalytic activity while domain D3 mainly acts as an allosteric effector increasing the rate of reaction (Fedorova, 2012). Domain D2 and D4 have been shortened to stem loops, and domain D6 was removed entirely. As a ribozyme, D135 cleaves its exonic substrate under high concentrations of salt with multiple turnovers. Because domain D6 is lacking, this cleavage is triggered by a water molecule, which attacks the exposed 5'-splice site as a nucleophile (Swisher *et al.*, 2001). The fact, that cleavage is taking place at all, implies a correctly folded RNA. To study the folding of D135 by smFRET two artificial loops serving as labeling platforms were introduced into domain D1 and D4 (D135-L14) (Steiner, 2008). Further, the 3'-end of D135 was elongated to enable the classical surface immobilization via hybridization to a biotin-carrying DNA oligonucleotide. The previously developed construct D135-L14 serves as a starting point for the present study, to which we compare the obtained findings. smFRET studies of the engineered construct D135-L14 revealed four folding states; two extended intermediates I_1 at $FRET_{I1} = 0.1$ and I_2 at $FRET_{I2} = 0.2$, which were originally defined as one intermediate state I (Steiner, 2008). Furthermore, a folded intermediate F at $FRET_F = 0.4$ and the hypothesized native state N at $FRET_N = 0.6$ were detected (Steiner, 2008; Steiner *et al.*, 2009; Karunatilaka *et al.*, 2010; Fiorini *et al.*, 2015). All observed FRET states are in the lower FRET range and strongly overlap. It was shown that the proposed native state is favored under certain conditions, e.g. in the presence of the co-factor Mss116 or the exon-based substrate (17/7) consisting of 17 nt of the 5'exons followed by the first 7 nt of the intron. D135-L14 is supposed to fold stepwise and directly into a compact native structure (Steiner, 2008). Intramolecular tertiary interactions are involved in folding and are essential to yield a functional, thus, catalytically active RNA. The individual domains of the group IIB intron *Sc*.ai5γ form long-range tertiary interaction between themselves and with its flanking exons (FIGURE 1.4). In particular, those between the intron and its exons are known to be essential for catalysis (Costa *et al.*, 2000; Qin and Pyle, 1999). These tertiary contacts are formed between two intron-binding sites (IBS1/2) located on the 5'exon and one intron-binding site (IBS3) on the 3'exon and their particular counterparts, the exon-binding sites (EBS1-3), on the intron. The first time, the influence of the exons was considered for splicing studies in 2010 when Zingler and co-authors designed the wild-type construct *Sc*.ai5γ with short (20 nt) and long (300 nt) flanking exons and investigated the contribution of the co-factor by cleavage assays (Zingler *et al.*, 2010). Their research focus was the splicing mechanism under near-physiological conditions; low salt concentration, in the presence of the co-factor, and an intron with flanking exons (Zingler *et al.*, 2010). It was only in 2012 that Pyle and co-workers described the importance of the flanking exons for the RNA fold in 2012 (Fedorova, 2012). NAIM (Nucleotide Analog Interference Mapping) indicates the formation of the IBS2-EBS2 duplex to play a major role in RNA compaction, stabilization, and global folding. Until today, the importance of the exonic binding sites and their contribution to the folding and function of the ribozyme are not fully appreciated. Thus, we refined the construct D135-L14 by adding the flanking exons to the termini (eD135e-L14), as they are integral to understand the folding and activity of the intron.

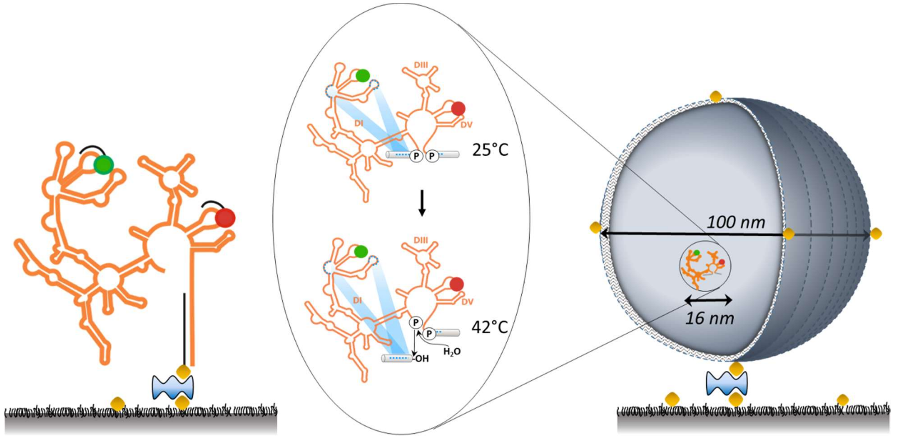


Figure 2.13 Classical surface immobilization of D135-L14 (left) versus vesicle encapsulation of eD135e-L14 (right). smFRET studies on the folding of the intron derivative D135-L14 were carried out on surface-tethered RNA molecules. The RNA is immobilized via a biotinylated (yellow square) DNA oligonucleotide on a streptavidin-coated surface. eD135e-L14 is encapsulated into phospholipid vesicles which are anchored in the same way. The encapsulation enables to follow folding and cleavage of the 5'exon of an active ribozyme at 42 °C on free diffusing molecules. Figure was adapted from (Steiner, 2008).

We are interested in a more detailed picture how precisely the exons influence the RNA folding and activity. Because the flanking exons are directly involved in catalysis at getting excised, the classical surface immobilization via the exons is not possible for eD135e-L14. Thus, we adapted the encapsulation approach for smFRET to a functional intron derivative eD135e-L14 to study the dynamics of a freely diffusing ribozyme and the role of its flanking exons. In this way, the RNA is not hindered in its motion and interactions with the surface can be neglected. *In vitro*, the cleavage of the 5'exon is induced by high concentrations of salt (500 mM KCl, > 40 mM $MgCl_2$) and elevated temperature (42 °C) and results in an excised 5'exon and a shortened precursor D135e-L14. At 25 °C, all derivatives of group II introns are inactive. To be able to observe an active RNA throughout its reaction, we had to induce RNA splicing on the smFRET microscope inside the immobilized vesicles by increasing the temperature to 42 °C. We are showing a comparison between the classical surface immobilization and the liposome encapsulation of D135-L14 and study the influence of both approaches on the folding of this already well-characterized construct. In addition, we applied stroboscopic alternating laser excitation (sALEX) to detect fast conformational changes on the low microsecond timescale (Kapanidis *et al.*, 2005; Kapanidis *et al.*, 2015). By this, we are combining the benefits of a high time resolution on freely diffusing molecules but captured in surface-immobilized vesicles.

2.2 Results and Discussion

2.2.1 Classical surface-immobilization versus novel RNA encapsulation

Following RNA folding and cleavage of eD135e-L14 by single-molecule FRET is only possible if the RNA is freely diffusing and not tethered via the exons to a surface. Both flanking exons are directly involved in catalysis and need to be free to interact with the intron. By encapsulating eD135e-L14 into phospholipid vesicles and inducing folding on the microscope, we capture all splicing products. In order to achieve this goal, first, we had to adapt the experimental procedure of encapsulation to our model system. Therefore, protocols from Cisse et al. and Liu et al. were optimized, and every deviation from the original procedure was confirmed by ensemble or single-molecule experiments (Cisse *et al.*, 2007; Liu *et al.*, 2010). For example, the optimal ratio of RNA to phospholipids to obtain single-molecule encapsulation was determined by single-step photobleaching (FIGURE 2.31), or the optimal temperature during encapsulation to prevent early splicing was investigated by ensemble native PAGE studies (FIGURE 2.17) (Cisse *et al.*, 2007; Liu *et al.*, 2010; Hadzic, 2017). After successful optimization of the encapsulation protocol, we investigated the impact of the two different immobilization techniques to the RNA fold. The starting point of our investigation was the classical surface immobilization of D135-L14 to a PEGylated quartz-surface. To elucidate the effect of surface immobilization versus encapsulation on the RNA folding, D135-L14 was encapsulated into phospholipid vesicles, both, in the presence and absence of its 3'-linker. The 3'-linker is a 23 nt long DNA oligonucleotide which hybridizes to the 3'-end of D135-L14 and is used to tether the RNA to a surface. In either case, the RNA was labeled fluorescently by DNA oligonucleotides in the loops L1 and L4. The next paragraph gives an overview over all molecules (cumulative histogram and representative FRET trace) and their composition (zero, static, and dynamic molecules) and shows a comparison of surface-attached D135-L14 versus its encapsulated version including the 3'-linker (FIGURE 2.14).

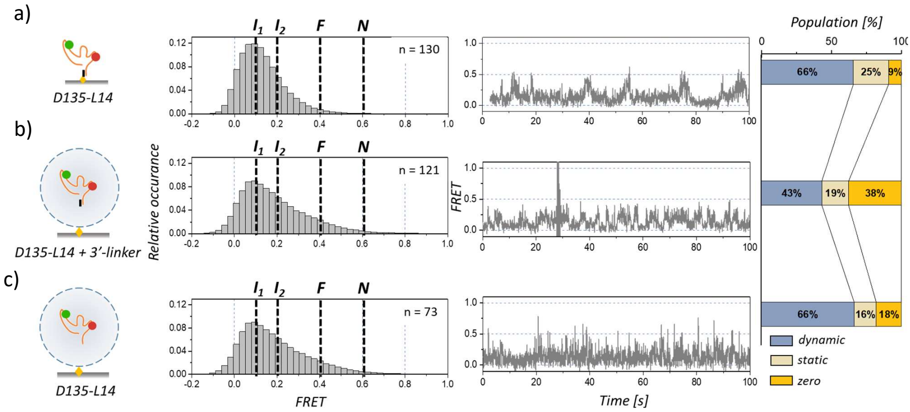


Figure 2.14 Influence of different immobilization techniques on the RNA folding of D135-L14 at 25 °C. A comparison of D135-L14 between ***a)*** *surface immobilized,* ***b)*** *encapsulated in vesicles and hybridized to the 3'-linker, and* ***c)*** *encapsulated without the 3'-linker is shown. Only static and dynamic traces were used for the FRET histograms. The folding states I_1 (E = 0.1), I_2 (E = 0.2), F (E = 0.4), and N (E = 0.6) are indicated in yellow. A typical trajectory showing the apparent FRET over time and the distribution of dynamic, static and zero FRET molecules is shown. The exposure time was 28 ms.*

Surface immobilization of D135-L14 was the technique used in the past to follow the RNA folding by smFRET. Under high salt concentrations, 500 mM KCl and 100 mM $MgCl_2$, we observe four FRET states as known from literature (Karunatilaka *et al.*, 2010; Fiorini *et al.*, 2015) (Fiorini *et al*. in preparation, 2018). Thereby, the 0.1 FRET state is the most populated state, followed by 0.2, 0.4 and very rarely occurring 0.6 (FIGURE 2.14). A comparison between all three conditions reveals no additional FRET states and a similar distribution of all subpopulations. The FRET histograms differ only in the shape of the tail. In the presence of the 3'-linker, the higher FRET states 0.4 and 0.6 become more populated than without. To understand whether this tailing is due to the presence of the linker or an encapsulation effect, D135-L14 was encapsulated in the absence of the 3'-linker (FIGURE 2.14). The shape of the FRET histogram looks the same, but a typical FRET trace shows higher dynamics than the encapsulated D135-L14 without the 3'-linker and D135-L14 surface immobilized. Not only the individual molecules show faster transitions, but also the number of dynamic molecules increases by ~23 %.

Next, the FRET trajectories were categorized into dynamic, static and zero FRET molecules. The criteria are for dynamic molecules at least one apparent transition between two FRET states, for static molecules a static FRET signal over the observation time, and for zero FRET molecules the presence of both dyes and a static zero FRET signal. Zero FRET refers to a distance between the two fluorophores larger than 10 nm. In the context of the RNA folding pathway, molecules showing zero FRET can be partially unfolded structures in which the two loops L1 and L4 are far apart. The percentage of all measured molecules showing zero FRET differs significantly for all conditions (FIGURE 2.14, APPENDIX CHAPTER II). Thus, for a general comparison between different conditions, we always consider molecules in a dynamic and static FRET state, and only the overall proportion of zero FRET molecules is shown in the column diagrams and summarized in APPENDIX CHAPTER II (FIGURE 2.14). A static molecule stays entrapped in one conformation within the observation window while a dynamic molecule switches between at least two. In order to show the influence of the individual subclasses, we compared these FRET histograms with the cumulative FRET histogram of all molecules in FIGURE 2.15. For all conditions, the different FRET states are equally represented in the histograms of the subclasses and do not significantly influence the overall FRET histogram. Except for zero FRET molecules, which are shifting the whole profile towards lower FRET values.

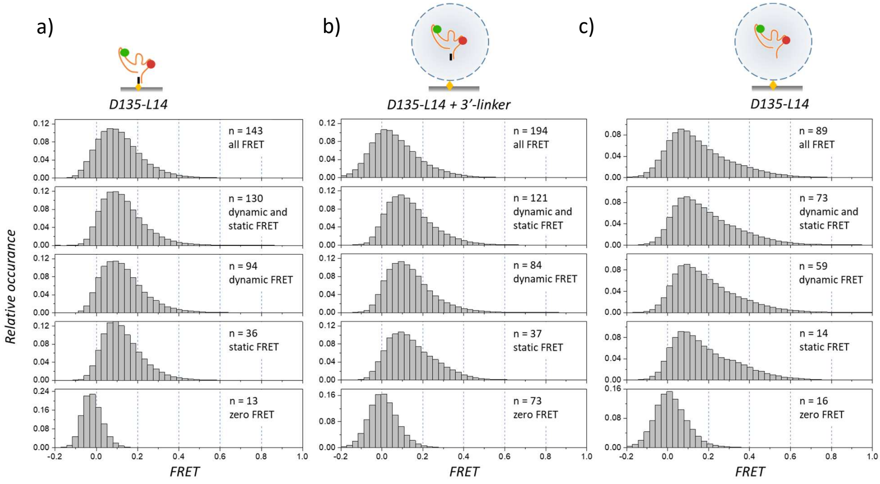


*Figure 2.15 Decomposing FRET histograms: the impact of molecules, which are dynamic, static, or showing zero FRET only to the overall FRET histogram. The different conditions are **a)** D135-L14 surface immobilized, **b)** D135-L14 encapsulated in the presence of the 3'-linker and **c)** D135-L14 encapsulated without 3'-linker. All experiments were performed at 25 °C and with a time resolution of 28 ms. The respective numbers to build a histogram are indicated. Flanking exons stabilize two RNA folding states*

The influence of the flanking exons is shown on the comparison between the encapsulated D135-L14 and eD135e-L14. The ribozyme D135-L14 contains an elongated 3'-end, which was designed to attach the RNA to a surface. Our results demonstrated that a free single-stranded 3'-end induces artificial dynamics. In the context of investigating the influence of different immobilization techniques to the folding pathway, artificially introduced dynamics are not desirable. For that reason, we used the encapsulated D135-L14 with a stabilized 3'-end through the 3'-linker and compare it with the encapsulated eD135e-L14. Our results show that the FRET histograms of both constructs differ from one another (FIGURE 2.16). In contrast to D135-L14, two pronounced extended intermediate states I_1 and I_2 are detected for eD135e-L14. I_1 is present for both constructs but much less occupied for the exon-carrying ribozyme. Here, the second intermediate state I_2 is the most dominant population. The folded state *F* is also higher populated for eD135e-L14 in comparison to D135-L14, and native state *N* is for both introns the lowest populated one.

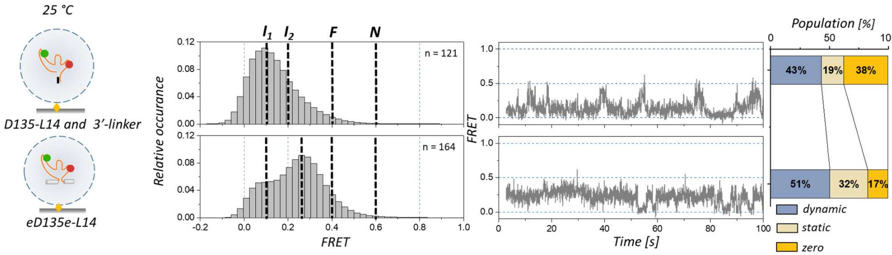


Figure 2.16 The influence of the flanking exons is shown regarding a comparison between D135-L14 in the presence of the 3'-linker and eD135e-L14, both freely diffusing in vesicles. The FRET histogram of eD135e-L14 shows two distinct FRET populations, I_1 at 0.1 and I_2 at 0.25, also reflected in the representative FRET trace. The higher FRET states F and N are less populated for both ribozymes. Only dynamic and static molecules are cumulated in the FRET histograms. The diagram on the left shows the distribution of dynamic, static and zero FRET molecules. All experiments were performed at 25 °C and 28 ms.

2.2.2 Temperature rise induces ribozyme activation

Preliminary ensemble experiments using native Polyacrylamide gel electrophoresis (PAGE) were required to define the optimal temperature for smFRET experiments under which splicing occurs. For investigating the splicing mechanism, the labeling positions were changed from internal loops to the flanking exons. Because the flanking exons are short and the intron-binding sites have to remain unassigned, short peptide nucleic acids (PNAs) were chosen as labels. This way, the ribozyme activity can be followed over time for eD135e-L14 at different temperatures (FIGURE 2.17C). The results showed that the cleavage of the 5'exon requires not only require high salt concentrations but elevated temperatures of 42 °C.

The stabilization effect of the flanking exons to the intron fold also becomes visible on native PAGE (FIGURE 2.17). D135-L14 and eD135e-L14 were fluorescently labeled at the same positions internally by DNAs (FIGURE 2.17A) or PNAs (FIGURE 2.17B). The impact of the different labels is topic of PARAGRAPH 2.2.4. In each case, three main bands are visible from top to bottom, an unbound PNA fraction, an intermediate, the primary precursor RNA and an unbound DNA fraction. A comparison of the intermediate and precursor fraction between D135-L14 and eD135e-L14 shows a smear in the absence of the flanking exon and distinct, well-separated bands in the presence of the flanking exons.

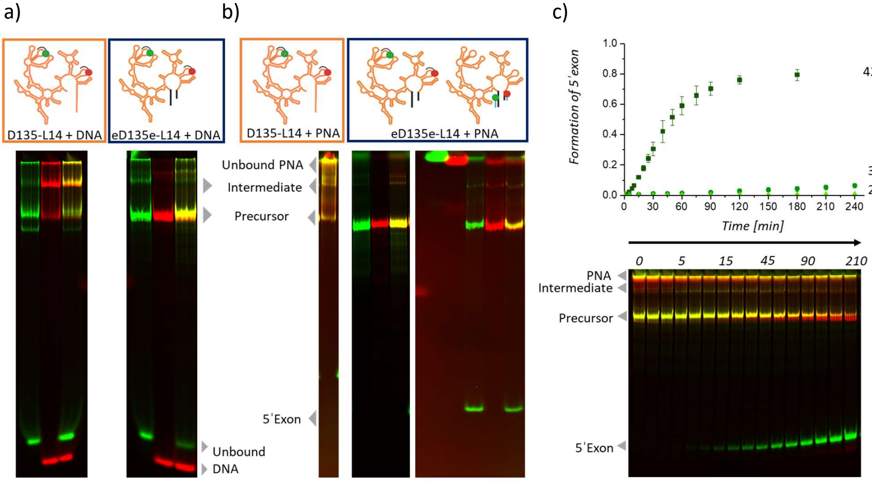


Figure 2.17 Native PAGE studies of D135-L14 and eD135e-L14. Shown is an overlay of the Cy3 and the Cy5 channel. A co-localization of Cy3 (green) and Cy5 (red) results in yellow. ***a)*** *Both ribozymes, D135-L14 and eD135e-L14, are fluorescently labeled with DNA labels in L1 and L4, starting with only DNA-Cy3 label, only DNA-Cy5 label, and both labels.* ***b)*** *A comparison between D135-L14 and eD135e-L14 labeled with PNAs is shown. PNAs are not migrating through a gel because they are not charged. eD135e-L14 can be labeled additionally at the flanking exons by PNAs.* ***c)*** *Like this, the cleavage of the 5'exon (green) can be followed over time. The ribozyme eD1353-L14 is only active at 42 °C.*

For observation an active ribozyme on a single-molecule level, we induced the first step of splicing on the microscope by increasing the temperature of the objective and the sample holder to 42 °C and followed the RNA folding over 2 h. After this, we were cooling down the chamber to 25 °C and studied changes before and after RNA activation (FIGURE 2.18).

Elevated temperature destroy surface contact of D135-L14, and thus the ribozyme cannot be activated. Instead, it has to be encapsulated to investigate the folding at 42 °C. Independent whether the 3'-linker is hybridized to D135-L14 or not, an increase in temperature does not affect the FRET histograms or individual trajectories but leads to a decrease in the number of dynamic molecules by about ~20 % (FIGURE 2.18A). After an incubation time of 2 h at 42 °C, the RNA was cooled down to 25 °C. In comparison with the starting temperature and 42 °C, the FRET histograms for both conditions look unchanged, only the number of dynamic molecules is changing. The number of dynamic molecules remain low after cooling down to 25 °C for D135-L14 with a flexible 3'-end, this indicate a structural change, which stabilizes a particular RNA conformation.

While the temperature does not affect the FRET histograms for D135-L14, the histograms for the exon-carrying intron changes. By raising the temperature for eD135e-L14, an increasing number of state transitions indicating higher conformational dynamics become visible in the FRET histogram and the representative traces (FIGURE 2.18C). The two central FRET populations at ~0.1 and ~0.25 are still represented, but their maxima are closer together and less separated at 42 °C. The transition between these FRET states become faster, which leads to an overlap of FRET states. The fraction of static molecules is decreasing to favour of the dynamic molecules (FIGURE 2.18C). The relative occurrence of zero FRET varies between different conditions (FIGURE 2.18, APPENDIX CHAPTER II). The population of zero FRET molecules of eD135e-L14, which represents an unfolded species increases upon temperature rise and returns to the initial situation after cooling down to 25 °C (FIGURE 2.18C).

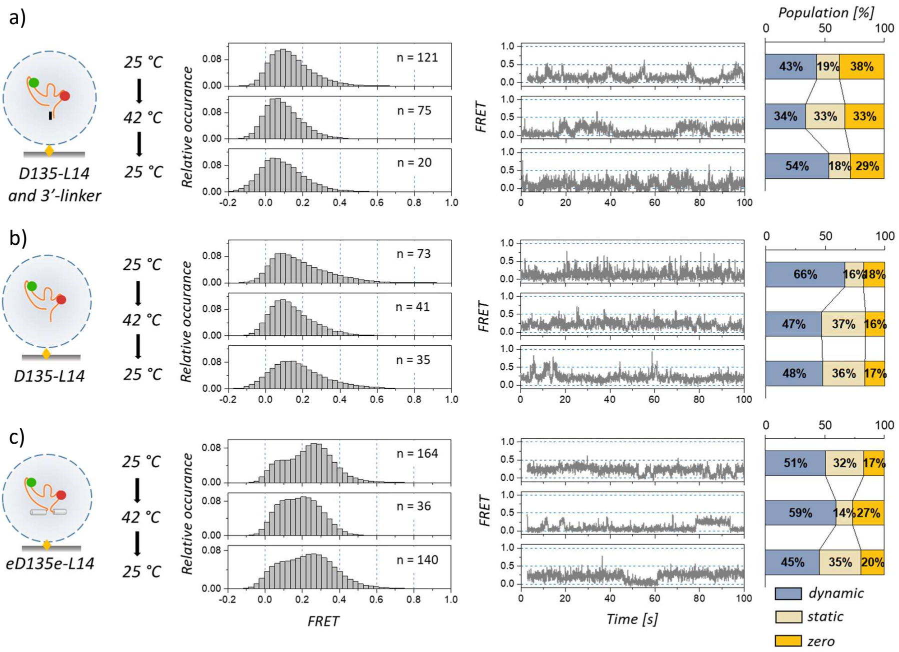


*Figure 2.18 Influence of different temperatures on the folding of the encapsulated constructs **a)** D135-L14 with and **b)** without the 3'-linker and **c)** eD135e-L14. Shown are the result for 25 °C (inactive ribozyme), 42 °C (active ribozyme) and cool down to 25 °C after a 2 h incubation at 42 °C (cleaved ribozyme). Only dynamic and static molecules were taken for the FRET histograms. The diagrams on the left show the distribution of dynamic, static and zero FRET molecules. The exposure time was 28 ms.*

A closer look to the individual subclasses of dynamic, static and zero FRET molecules shows that independent of the temperature, for D135-L14 all three populations show identical FRET histograms, and no trend is visible (FIGURE 2.19). In contrast to D135-L14 with the 3'-linker, a separation of FRET histograms into the individual subclasses for eD135e-L14 shows differences in FRET distribution (FIGURE 2.20). The static FRET molecules are more populated in the expanded intermediate state at ~0.25 than the dynamic molecules are. The dynamic molecules are occupying both intermediate states equally for all temperatures. The zero FRET molecules are influencing the FRET histogram including all molecules by shifting the ~0.1 FRET population towards lower FRET values.

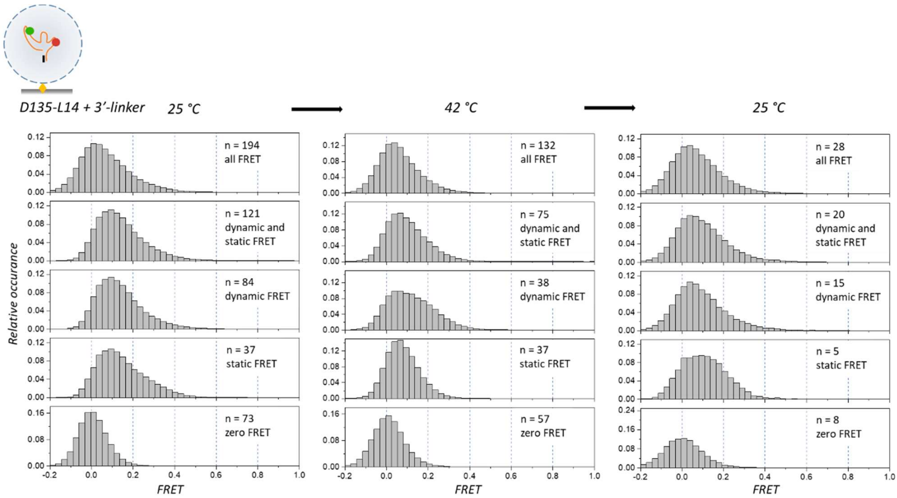


Figure 2.19 The impact of the three subclasses of dynamic, static and zero FRET states to the overall FRET histogram for the encapsulated D135-L14 in the presence of the 3'-linker is depicted. The influence of different temperatures (inactive, active and cleaved ribozyme) on the individual subclasses is shown. The exposure time was 28 ms.

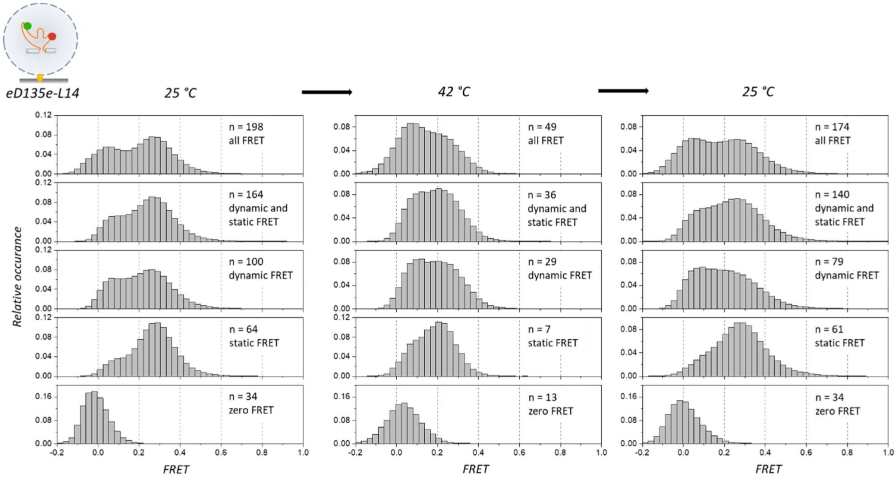


Figure 2.20 The impact of the three subclasses of dynamic, static and zero FRET states to the overall FRET histogram for the encapsulated construct eD135e-L14. The influence of different temperatures (inactive, active and cleaved ribozyme) to the subclasses is shown. The static molecules predominantly populate the extended intermediate state at ~0.25, while the dynamic molecules are occupying both intermediate states equally. The exposure time was 28 ms.

For a comparison between smFRET trajectories recorded at different temperatures, the change in photophysical properties of the dyes have to be considered since it influences the FRET efficiency. The observed FRET efficiency depends on the quantum yield of both fluorophores and their detection efficiency (McCann *et al.*, 2010; Sanborn *et al.*, 2007). The quantum yield Φ of the free dyes Cy3 and Cy5 was determined experimentally (Steffen, 2015) and is strongly temperature dependent (at 25 °C $\Phi_{Cy3} = 0.31$, $\Phi_{Cy5} = 0.07$; at 42 °C $\Phi_{Cy3} = 0.22$, $\Phi_{Cy3} = 0.05$). The temperature dependence of the quantum yield of both dyes is linear. The detection efficiency of the camera η depends on the set of filter used on the smFRET setup ($\eta_A = 0.91$, $\eta_D = 0.95$). In general, a correction of the FRET efficiency by applying the gamma factor γ is important for statements about absolute FRET efficiencies between different conditions but is negligible for a relative comparison (McCann *et al.*, 2010). Applying the gamma factor according to EQUATION 2.3 results in high corrections factors $\gamma_{25°C} = 4.24$ and $\gamma_{42°C} = 4.21$, but the differences between both temperatures are relatively small. We did not determine the quantum yield for both dyes linked to the DNA label and hybridized to the RNA. The surrounding of the fluorophores strongly influences their quantum yields, and thus applying the quantum yields of the free dyes yields aberrant gamma values. Hence, they are not taken into account for the analysis.

$$\gamma = \frac{\eta_A}{\eta_D} * \frac{\Phi_A}{\Phi_D} \qquad \textit{Equation 2.3}$$

We are aware of photophysical effects cause by RNA-dye interactions and recently discovered as an RNA-induced fluorescence enhancement (RIFE), but we are not correcting for it in this study (Steffen *et al.*, 2016).

2.2.3 High time resolution

RNA folding takes place on a timescale of minutes for a global fold to milliseconds for local collapses (Russell *et al.*, 2002). The achieved time resolution is limited by the frame rate of the electron multiplying charge-couple device (EMCCD). Our standard conditions were an exposure time of 28 ms while applying alternating laser excitation (ALEX) and a frame rate of 4000. Higher time resolutions enable to visualize dynamics below our detection rate. According to experiments from Farooq and Holbein applying a stroboscopic illumination (sALEX) leads to FRET distributions, which were otherwise averaged out (Farooq and Hohlbein, 2015). To study whether additional FRET states or dynamics are hidden by the time resolution of 100 ms or 28 ms, we applied higher time resolutions of 14 ms and 2 ms. For latter, the detection rate of the EMCCD was 14 ms, but the laser excitation pulse was set to 1.7 ms (sALEX). The laser power was adjusted according to the exposure time to keep the number of detected photons constant (TABLE 2.7). By increasing the laser power, the fluorophores become unstable. The number of photobleaching events was increasing drastically. FIGURE 2.21 shows FRET histograms of different exposure times. For a better comparison, the data of 28 ms, 14 ms, and 2 ms were binned to 100 ms (black line). The binned data always show two main populations. Thereby the ~0.25 FRET state is, in all cases, the most populated one. A comparison between the different exposure times shows that from 100 ms towards 14 ms, the FRET histogram become narrow and at the same time the maxima of the two main FRET states at ~0.1 and ~0.25 get closer (FIGURE 2.21). For the unbinned data, at a time resolution of 2 ms, the FRET histogram is very broad, and the two main populations are not distinguishable here. The background signal is significantly higher at 2 ms, induced by the high laser power. In addition, the high laser power caused photophysical effect of the dyes, e.g. photoblinking. The number of dynamic, static, and zero FRET molecules remains the same overall different exposure times, which is evidence that no further dynamics at higher time resolution occurs.

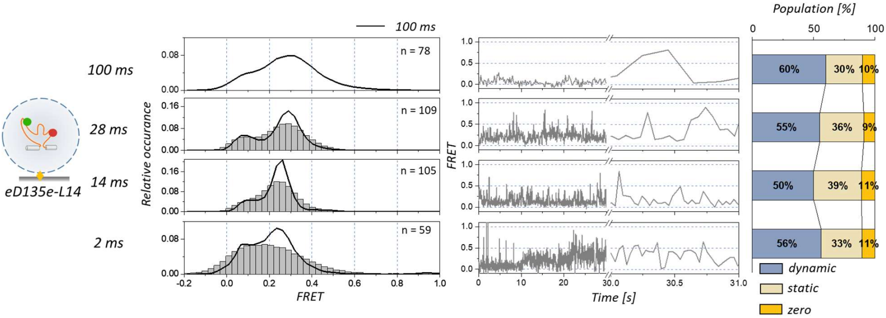


Figure 2.21 smFRET histograms of the encapsulated construct eD135e-L14 at different exposure times. The data for 28 ms, 14 ms, and 2 ms were binned to 100 ms (black line). The representative traces are showing the increase of dynamics with increasing time resolution, for 10 data points per second (100 ms), ~35 data points per second (28 ms), ~70 data points per second (14 ms), ~70 data points (1.7 ms laser pulse, but 14 ms camera detection). The number of dynamic, static and zero FRET molecules remain the same over all time resolutions. The experiments were carried out at 25 °C.

2.2.4 DNA labels compared to PNAs

PNAs are synthetic analogs of DNAs evolving as a promising alternative for labeling RNA. PNA labels have two main advantages over DNA labels; 1) PNA have a higher binding affinity to RNA (Robertson *et al.*, 2006; Muratovska *et al.*, 2001; Chinnery *et al.*, 1999). For that reason, shorter sequences can be used to label the RNA, minimizing the number of exogenous bases pairs. 2) A higher binding affinity yields a higher labeling efficiency by using the same amount and ratio of labels to RNA. Previously, a smFRET study on D135-L14 compared DNA and PNA labels with each other and determined no difference in RNA folding (Schmitz *et al.*, 2015). For this reason, also PNA labels were used to label eD135e-L14 internally in loop L1 and L4 (FIGURE 2.22) and later externally at the flanking exons (CHAPTER III).

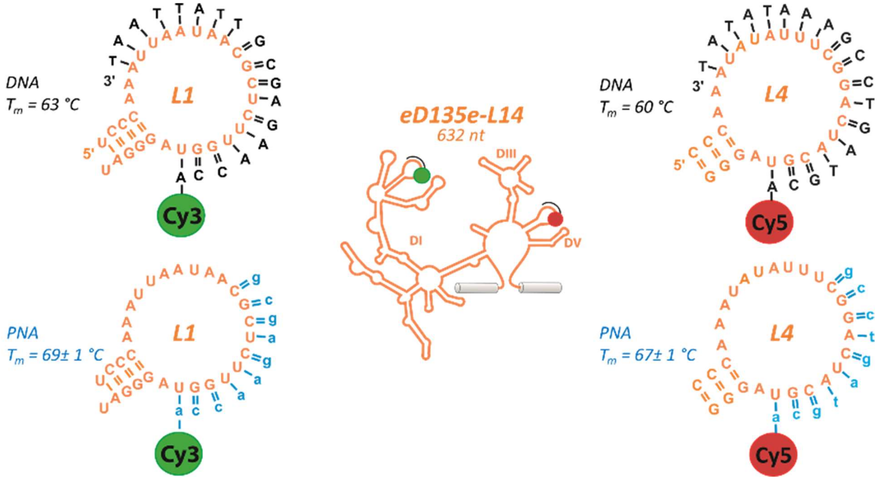


Figure 2.22 DNA (black) versus PNA (blue) probes for fluorescence labeling of eD135e-L14 (orange). The PNAs are shorter by eight bases than the DNAs, but the experimentally determined melting temperature of PNAs is still higher than for DNAs. (Fazliji, 2017; Schmitz et al., 2015).

Interestingly, depending on the fluorescent label for eD135e-L14 we observe differences in FRET histograms and distribution of dynamic, static and zero FRET molecules, unlike previously reported. The FRET histograms show less separated FRET states in the presence of the PNA labels (FIGURE 2.). An increase in temperature leads in both cases to a blurring of the states. The maxima of the expanded intermediates are coming closer together, indicating higher dynamics. The number of static molecules is significantly increased for PNAs upon raising the temperature (FIGURE 2.B). While the proportion of zero FRET remains unchanged, dynamic molecules become statics. Overall, the dynamics by applying PNA labels increase, which could result from the labeling platforms L1 and L4. PNAs are shorter and thus hybridize only to half of the loops, which give them additional flexibility.

Dynamic, static and zero FRET molecules show the same trends as the DNA labels (FIGURE 2.24). At both temperatures, the static molecules are more present in the expanded intermediate I_2 at 0.25 than in I_1 at 0.1. In contrast, the dynamic molecules are populating more the ~0.1 FRET state. The influence of zero FRET molecules is minimal for PNAs because only 8 % of all molecules are trapped in the zero FRET state.

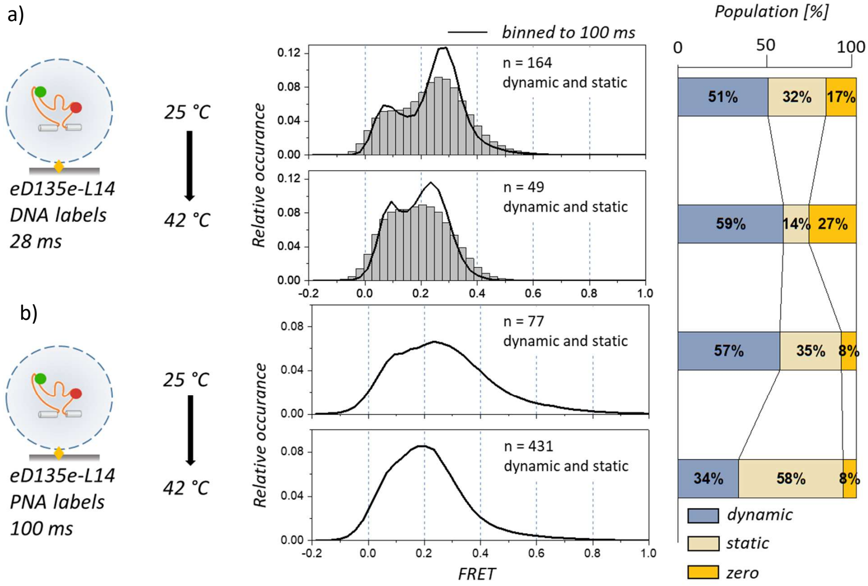


Figure 2.23 Influence of the fluorescent labels for the encapsulated eD135e-L14, a) DNA versus b) PNA labels. For a better comparison between both labels, the two upper FRET histograms for eD135e-L14 with DNA labels were binned to 100 ms (black line). Only dynamic and static molecules were cumulated for the FRET histograms. The column diagram is showing the proportions of dynamic, static and zero FRET molecules at different conditions.

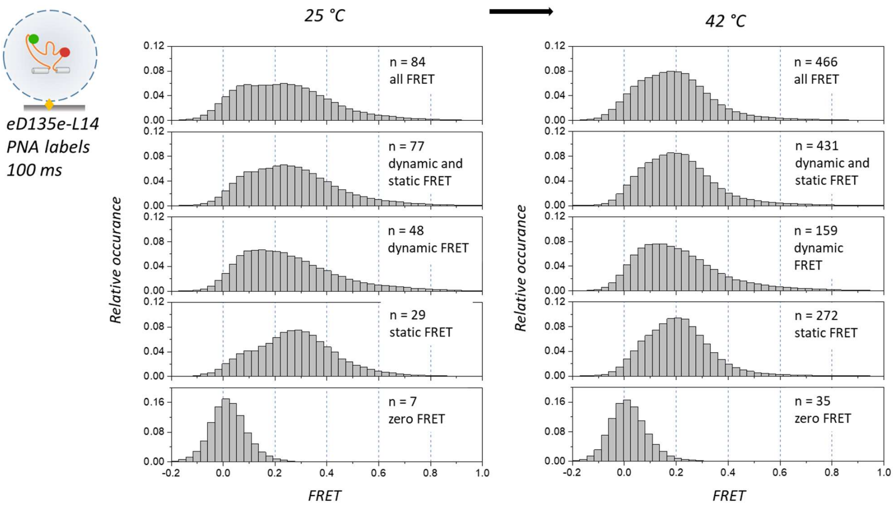


Figure 2.24 Molecules were sorted into dynamic, static and zero FRET molecules for the encapsulated eD135e-L14 labeled with PNAs at 25 °C and 42 °C. The exposure time was 100 ms. Static molecules are mainly occupying the ~0.25 FRET state, while dynamic molecules are mainly populating the ~0.1 FRET state.

We were interested in how shifts in the FRET efficiency are evolving over time and temperature for eD135e-L14 with PNA labels. Unfortunately, no trend in FRET histograms becomes visible. The maxima of both central populations at 0.1 and 0.25 are equally distributed. After an incubation time of 90 min, coalesce of both peaks becomes visible. This indicates the occurrence of structural rearrangements during ribozyme activation.

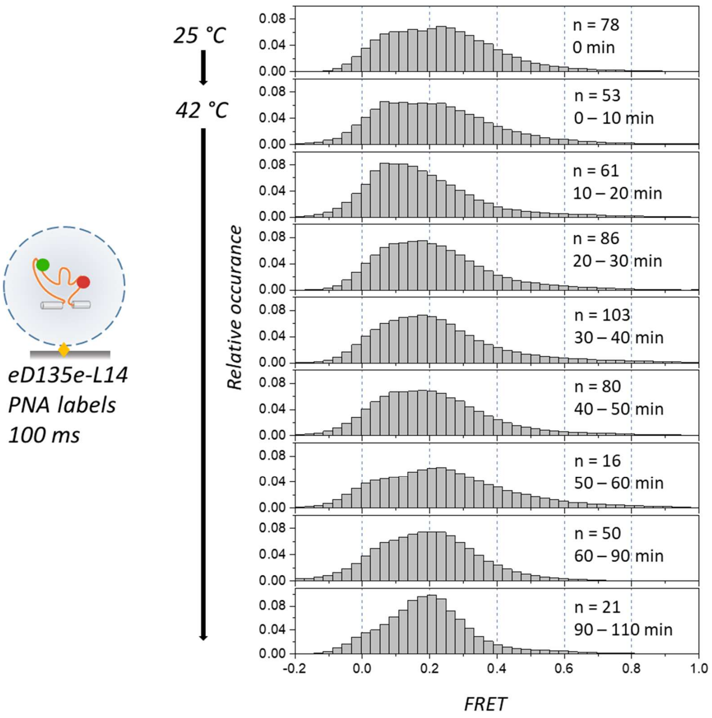


Figure 2.25 FRET histograms and how they are evolving over time and temperature for eD135e-L14 with PNA labels. The number of molecules is indicated on top of the incubation time at 42 °C. The exposure time was 100 ms.

2.2.5 Discussion

We hypothesize that the encapsulation alone is influencing the population of higher FRET states. Higher FRET states are associated with a more compact folded RNA. By anchoring the D135-L14 to a surface, the motion of the molecule is affected in a way that RNA folding is hampered. If D135-L14 can freely diffuse within a vesicle, the RNA can undergo folding without any hindrance. The presence of the 3'-linker is influencing the RNA dynamics. The 3'-linker and the intronic 3'-end are forming a stable DNA/RNA-helix of 23 nt. In the absence of the 3'-linker, D135-L14 has a 31-nucleotide long single-stranded RNA tail. The individual trajectories show higher dynamics and the overall number of dynamic molecules increases in the absence of the 3'-linker (FIGURE 2.14). The relative occurrence of dynamic and static molecules are compatible with data from previous studies in the case of D135-

L14 (70 % (Fiorini, 2015); 85 % (Hadzic, 2017); 50 % (Steiner, 2008)). The only exception is the publication of Krishanthi at al. where they observed only 25 % of dynamic molecules (Karunatilaka *et al.*, 2010). The difference between our values and the ones reported in the literature are attributed to different exclusion criteria and the use of automated single-molecule localization, which reduces the user-bias drastically.

Alternating laser excitation allows confirming the presence of both dyes, donor and acceptor, on the molecule under observation (Kapanidis *et al.*, 2005; Kapanidis *et al.*, 2015). In 2008, Steiner and co-authors could not distinguish between a zero FRET or a donor-only signal, because they applied only single donor-excitation (Steiner, 2008; Steiner *et al.*, 2009). Repeating previous experiments but making use of ALEX reveals molecules that carry both fluorophores but show a static zero FRET signal. Such zero FRET states of the surface-immobilized ribozyme D135-L14 can be attributed to unfolded or misfolded RNA molecules, which have donor and acceptor labels spaced more than 10 nm apart. Interestingly, the zero FRET state is also present in dynamic molecules, but only to a small percentage and with short dwell time, suggesting that the zero FRET state belongs to the overall folding pathway. For the encapsulated ribozyme D135-L14 in the presence of the 3'-linker, the number of molecules showing zero FRET merge into dynamic molecules upon temperature increase. Molecules, which evolve from zero to dynamic FRET over time, are evidence for misfolded molecules that fold and become active. The presence of an inactive and thus potentially misfolded fraction as a part of the ribozyme folding path was already reported many times (Daniels *et al.*, 1996; Su *et al.*, 2003; Zingler *et al.*, 2010; Zingler, 2014). Investigations by Su et al. on the refolding of D135 showed that only 67 % of the RNA molecules fold back into the active conformation (Su *et al.*, 2003). In this context, the observed 30 % of zero FRET molecules in this smFRET folding study, are most likely in a misfolded state. Regarding the encapsulated ribozymes, the zero FRET state can also refer to a misfolded intron or additionally indicate unbound encapsulated labels because they remain inside the vesicle and cannot be flushed away. The probability of encapsulating two unbound labels is very low since the labeling efficiency is very high (80 %). Dissociation of the labels from the RNA within the vesicle is not expected due to the high effective PNA concentration inside the vesicle.

The flanking exons primarily affect RNA folding. eD135e-L14 is stabilized in two main conformations by the exons, the extended intermediate I_2 at 0.25 is predominant. We found an increase of 25 % in static molecules for the construct with flanking exons. This stabilization of the RNA is mainly attributed to the 5'exon where two intron binding sites (IBS1 7 nt, IBS2 6 nt) are located and form important tertiary contacts. Upon IBS-EBS binding, the RNA is captured in a more compact fold. This observation confirms previous NAIM studies from Fedorova et al. about the major role of IBS-EBS formation in RNA stabilization (Fedorova and Pyle, 2012).

Even more, we show that the exons are capturing the RNA in an extended intermediate fold (I_2) and stabilizing the overall construct, leading to fewer dynamics. These findings indicate that the previously assigned extended intermediate I_2 represents a catalytic active state. The stabilization of an intermediate folding state is also visible in the native PAGE (FIGURE 2.17), where the precursor RNA of eD135e-L14 appears as a single distinct band. In contrast to D135-L14, the precursor RNA is smeared, indicating multiple intermediates and higher dynamics.

Comparing the results at 25 °C with those at 42 °C for eD135e-L14 shows that increasing the temperature is broadening the FRET states. The flexibility of the single domains d2b and D4 where the labels are hybridized as well as the overall dynamics of the RNA are increasing. Interestingly, D135-L14 in the presence or absence of the exons show an opposite distribution of dynamic and static molecules upon temperature rise. For D135-L14, an increase in temperature leads to a decrease in dynamic molecules. Usually, higher temperatures lead to an increase in intramolecular dynamics. We might argue that the dynamics are taking place on a shorter time scale than our exposure time and therefore average out at 28 ms. For eD135e-L14, elevated temperature yield an increase in dynamic molecules by 20 %.

Cooling down to 25 °C, is inactivating the ribozyme again, with the difference that now the 5'exon is cleaved. The splicing assays reveal 75 % of cleaved 5'exon after an incubation time of 2 h at 42 °C (FIGURE 2.17). In the smFRET results, we cannot distinguish between the chemical cleavage reaction and the dissociation of the 5'exon. Nevertheless, we have evidence that cleavage occurred (splicing assays), and the 5'exon remains bound to the intron afterwards, e.g. no increase of zero FRET states at 42 °C. Taking the following issues into account i) crystal structure of *O. iheyensis* D1 joined to its 5'exon (Zhao *et al.*, 2015) and NAIM studies (Fedorova and Pyle, 2012) and ii) the high effective RNA concentration per vesicle (about 3 µM) the probability that the cleaved 5'exon stays bound to the intron is very high. The fact that the FRET histograms before and after cleavage look similar indicates that the 5'exon stabilizes the intron independent whether it is covalently bound to the intron or cleaved off.

A higher time resolution usually leads to narrower FRET states. At 2 ms time resolution, the broadening of the FRET histogram can be explained by an increase in background signal and photophysical effects introduced by the high laser power. The fact that no additional FRET populations are appearing and that the number of dynamic and static molecules does not change with increasing time resolution indicates that no additional folding states are hidden by the standard exposure time of 28 ms.

The influence of different labels was investigated by comparing DNA oligonucleotides with PNAs. While DNAs are hybridizing to the whole loops L1 and L4 of the ribozyme and forming a stable DNA/RNA helix throughout the entire loop, PNAs are occupying only half of the loops (FIGURE 2.22). We hypothesize that the unbound sequence allows for more flexibility, which leads overall to more dynamics. An increase in temperature leads to a higher number of dynamic molecules for DNA labels, but the opposite effect is visible for PNA labels. Most likely, the dynamics for PNA labels are increasing that much, that they are not detectable anymore. We observe the same effect with the very dynamic molecule D135-L14. At higher temperature, interconversions between states are too fast to be captured by the camera.

Our comparison study of the D135 and eD135L14e indicates that the latter is stabilized by the flanking exons, suggesting a revised folding scheme (FIGURE 2.14). At the beginning of the splicing pathway, only secondary structure elements are formed. This fraction can be attributed to the unfolded state *U* in which the fluorescent dyes are far away from each other, and no FRET can take place (*U* corresponds to zero FRET). Further folding into the complex tertiary structure is induced by magnesium (II) and elevated temperature. It is well known from the literature that the first and rate-limiting folding step is the correct folding of domain D1 (Zhao *et al.*, 2015; Su *et al.*, 2005). The folding of D1 is directly linked to the binding of the 5'exon. For that reason, we propose that the first extended intermediate state at ~0.1 reflects the folding of D1 and subsequent binding of the 5'exon. If the 5'exon is correctly positioned hydrolytic cleavage at the 5'-splice site takes place. Splicing assays confirm the fast cleavage of the 5'exon. The framework of folded D1 bound to the 5'exon serves as a central contact for the remaining domains. In a second step, the 3'exon is brought into close proximity of the 5'-splice site through binding to EBS3. To position the 3'exon correctly, the remaining domains D2, D3, D4 are binding to D1 while D5 is flipped into the core of D1 thereby forming the active site (Zhao *et al.*, 2015). The next step of the splicing process is the cleavage of the 3'exon followed by the ligation of both exons. Due to the missing domain D6, exon ligation is not happening for the construct eD135e-L14. This could explain the rare occurrence of the compact folds *F* and *N*. The splicing is stalled in the *I₂* state because the splicing reaction cannot proceed. D135-L14, on the other hand, is mainly trapped in the *I₁* state because the flanking exons for stabilization and catalysis are missing.

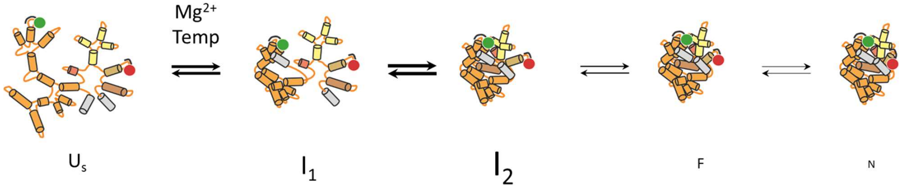


Figure 2.26 Proposed folding pathway of the freely diffusing intron eD135e-L14. The domains are colored in orange (D1), dark orange (truncated D2), yellow (D3), light brown (truncated D4) and brown (D5) and the flanking exons are depicted in grey. The unfolded state U_S refers to a RNA already assembled into secondary structure elements and fluorescently labeled (green, red). By the addition of Mg^{2+} and elevated temperature, the RNA starts to fold into a tertiary structure. I_1 refers to the first extended intermediate where D1 is mainly folded and binds the 5'exon. The most prominent state is the second extended intermediate I_2 where the structure becomes more compact and folds towards the folded F and native N state. The font size refers to the occurrence of FRET states in the smFRET experiments. Figure 2. adopted from (Steiner, 2008).

2.3 Conclusion and Outlook

We show by single-step photobleaching (PARAGRAPH 2.4.10) and fluorescence intensity histograms (APPENDIX CHAPTER II) the successful mono-encapsulation of a large ribozyme, which is fluorescently labeled. We only induce splicing and by this activate the ribozyme via increasing the temperature to 42°C, once the ribozyme is labeled and encapsulated. By increasing the temperature, we accelerate the first step of splicing comprised of RNA folding and subsequent 5'exon cleavage and investigate the influence of the covalently bound exons.

RNA encapsulation influences the overall fold. In particular, high FRET states are more populated, but the overall interconversions between the states are not affected. Because we observe a similar folding behavior of the same construct for both immobilization techniques, we assume that interactions between the RNA and the vesicle are negligible. Experiments of the encapsulated D135-L14 in the absence of a stabilized 3'-end lead to a significant increase in dynamics. The rising dynamics clearly result from the 3'-tail, which was added artificially and thus, influences the folding of the intron sequence. The 3'-end is single-stranded in the absence of the 3'-linker and thus very flexible. Our conclusions are that encapsulating the RNA increases the probability of compact folds. Ideally, a construct is needed where the 3'-end is not artificially elongated but ends at the 3'-splice site. Because the single-stranded 3'-end of D135-L14 is causing artificial dynamics, we are comparing D135-L14 in the presence of its 3'-linker with eD135e-L14 to investigate the influence of the flanking exons.

D135-L14 is a highly dynamic molecule adopting mainly the extended intermediates I_1 and I_2 and shows fast transitions between these FRET states. Adding the flanking exons to D135-L14 shows their considerable stabilize two RNA conformations. A comparison between D135-L14 and eD135e-L14 shows clearly the predominantly second extended intermediate at ~0.25 in the case of eD135e-L14. Via the tertiary contacts EBS/IBS, the exons are holding the RNA together and stabilize the folding states. The stabilization effect is also reflected in the increasing number of static molecules, which remain in the extended intermediate I_2 at ~0.25 FRET state.

Splicing assays showed that both intron derivatives D135-L14 (Fiorini, 2015) and eD135e-L14 are active only at 42 °C where cleavage of the substrate occurs. Activation of the surface-immobilized ribozyme D135-L14 by increasing the temperature is not possible because the RNA molecules are losing surface contacts. Therefore, we use vesicle encapsulation to follow the folding of the ribozyme while increasing the temperature to activate the RNA. Raising the temperature results in a decrease of dynamic molecules in case of the encapsulated construct D135-L14. Whether this is a real observation or a further increase in dynamics below the time resolution of 28 ms is taking place has to be confirmed by sALEX for 42 °C. By activating the exon-carrying ribozyme eD135e-L14 via elevated temperature, more dynamics are detectable, and static molecules turn into dynamic molecules. After cooling down to 25°C, most of the introns have cleaved off their 5'exon. Interestingly, we do not observe a dissociation of the 5'exon. Instead, it stays bound to the intron due to the high effective RNA concentration inside the vesicle. Our results show that independent whether the 5'exon is covalently bound to the intron or cleaved, the stabilization of the two primary populations remains unchanged.

Higher time resolution revealed no additional populations and dynamics, confirming that RNA dynamics are taking place on a millisecond timescale. A higher time resolution leads to narrower FRET histograms. Raising the time resolution to 2 ms (sALEX) increases the readout noise which overlays the narrow FRET histogram and thus does not yield additional information. Additional measurements with a higher time resolution at 42 °C are needed to rule out faster RNA dynamics induced by higher temperature.

Depending on the type of label, DNAs or PNAs, we observe differences in FRET histograms and in the number of dynamic molecules. If the RNA (eD135e-L14) is fluorescently labeled by PNAs dynamics are increasing at 25 °C compared to DNA labels. In the presence of the much shorter PNAs, only half of the loops are hybridized to the PNAs and form a stable helix. For an ideal comparison between DNAs and PNAs, a new construct would need to be designed with shorter loops that are adjusted for the length of the PNA labels. It has been already shown that PNAs are very suitable for RNA labeling and studying RNA folding (Schmitz *et al.*, 2015). The labeling efficiency is higher for PNAs than for DNAs, which is very important for encapsulating fluorescently labeled RNA. The probability of encapsulating free dyes is decreasing drastically with PNAs.

Encapsulation will shape the future of smFRET studies for investigations on functional nucleic acids. We showed that visualizing RNA folding via encapsulated molecules allows following natural RNA motions without any disturbance caused by anchoring the RNA itself. In addition, the temperature-induced activation of the cleavage is only possible because of encapsulation, where the reactants are retained in the vesicles. Regarding the group II intron ribozyme, we showed that the flanking exons are essential for the RNA folding and considerably stabilize the extended intermediate state.

2.4 Materials and Methods

2.4.1 Buffers and kits

Table 2.2 Overview of all buffers and their composition, which were used to perform all experiments of this chapter.

Solution name	Composition	Ref.
Acrylamide gel (denaturing)	7 M Urea, TBE buffer, 8 % Acrylamide, 10 % Amonium persulfate, TEMED (Tetramethylethylenediamine)	
Acrylamide gel (native)	5 % Acrylamide, native gel buffer	
TAE buffer	40 mM Tris-base (pH 8.3), 40 mM Acetic acid, 1 mM EDTA	
TBE buffer	89 mM Tris-borate, 2 mM Na_2EDTA (pH 8.3)	
Transcription buffer	40 mM Tris-HCl (pH 7.5), 40 mM DTT, 2 mM Spermidine	(Gallo *et al.*, 2005)
LB medium	25 g/L LB	
LB agar plates	25 g/L LB, 15 g/L Agarose, 100 µg/mL Ampicillin	
Chemical competent cells buffer TBFI	30 mM KOAc , 100 mM RbCl , 50 mM $MnCl_2$, 10 mM $CaCl_2$, pH 5.8 with HCl	
Chemical competent cells buffer TBFII	10 mM MOPS, 10 mM RbCl, 75 mM $CaCl_2$, pH 7.0, 15 % Glycerol	
Native gel buffer	66 mM HEPES, 34 mM Tris (pH 7.4)	
Urea Loading Buffer	12 M Urea, 40 mM Tris-HCl (pH 7.5), 0.1 % Xyanole cyanide , 0.1 % Bromephenol blue, 230 mM sucrose, 0.8 mM EDTA (pH 8.0)	
smFRET experiments		
Anti-blinking buffer (AB)	0.5 M KCl, 80 mM MOPS (pH 6.9), 0.1 M $MgCl_2$, 3 mg/mL Trolox	(Ha, 2001; Cordes *et al.*, 2009)
Imaging buffer (IB)	SAB, 21.7 U catalase, 1.65 U glucose oxidase	(Ha, 2001)

Lipid composition	9.8 mg/mL DMPC, 2.1 mg/mL biotinyl cap PE, chloroforme	(Cisse *et al.*, 2007)
PEG composition	80 mg MPEG-SVA-5000, 8 mg Biotin-PEG-SVA-5000, 0.1 M Sodium carbonate	(Chandradoss *et al.*, 2014)
Standard buffer	0.5 M KCl , 80 mM MOPS (pH 6.9)	(Pyle and Green, 1994)
Sugar anti-blinking buffer (SAB)	AB buffer, 555 mM D-glucose	(Ha, 2001)
T50 buffer	50 mM NaCl , 50 mM TRIS (pH 7.6)	(Zhao and Rueda, 2009)

2.4.2 Molecular cloning

The intronic sequence of both constructs, D135-L14 and eD135e-L14 is identical from the 5'- to the 3'-splice site. D135-L14 contains an artificial elongation right after the 3'-splice site of 31 nucleotides, which serves for hybridizing a complementary biotinylated DNA oligonucleotide. Like this, the RNA can be attached to e.g. a streptavidin-coated surface. In contrast, eD135e-L14 contains the original exon sequences on both sides, 31 nt at the 5'-end and 14 nt at the 3'-end.

The starting point for cloning the construct eD135e-L14 was the plasmid pT7D135L14 generated by Miriam Steiner (Steiner, 2008). Two consecutive PCRs were performed. The 31-nucleotide long 5'exon and the 14-nucleotide long 3'exon with the restriction site for *Hin*dIII additionally were introduced in the first PCR by using the oligonucleotides oSP1 and oSP2 (TABLE 2.3, FIGURE 2.27). Both primers were overlapping the template DNA by 19 nt (oSP1), respectively 20 nt (oSP2). After applying the standard PCR conditions, the PCR product of the size of 650 bp was visualized on a 1 % agarose. The PCR was carried out using the following final concentrations: 200 µM of each dNTP, 0.5 µM of each primer, 0.5 ng of template DNA and 0.01 U of *Phu* polymerase (Thermo Scientific) and the corresponding buffer in a 50 µL reaction mixture. The PCR method of 27 cycles included an initial denaturation time of 15 s at 98 °C, followed by a primer annealing period of 30 s at 60 °C and an elongation period of 15 s at 72 °C. A final elongation of about 5 min at 72 °C was the last step. The PCR product was extracted from an agarose gel by the PCR clean-up kit (Wizard® SV Gel and PCR Clean-Up System, Promega) according to the manual. The purified 659 nt long PCR product was used as a template for the second PCR which was amplified with the primer pair oSP3 which consists of the T7 promoter and an *Eco*RI restriction site and oSP4 (TABLE 2.3). Only the primer annealing temperature was adopted to 55 °C; all other conditions were identical to the first PCR. The correct size of both PCR products was confirmed on an agarose gel (FIGURE 2.27).

Table 2.3 PCR primer and the corresponding sequence, length, and melting temperature are shown.

Primer name	5'- sequence	Length [nt]	Tm [°C]
oSP1	GGGAGACTTACTACGTGGTGGGACATTTTCGAGCGGTCTGAAAGTTATC	49	57
oSP2	CGAGAAGCTTGATAATACATAGTATCCAGGAACCGTACGTGCG	43	70
oSP3	GGCGAATTCGTGAATTGTAATACGACTCACTATAGGGAGACTTACTACGTGG	52	55
oSP4	CGAGAAGCTTGATAATACATAG	22	56

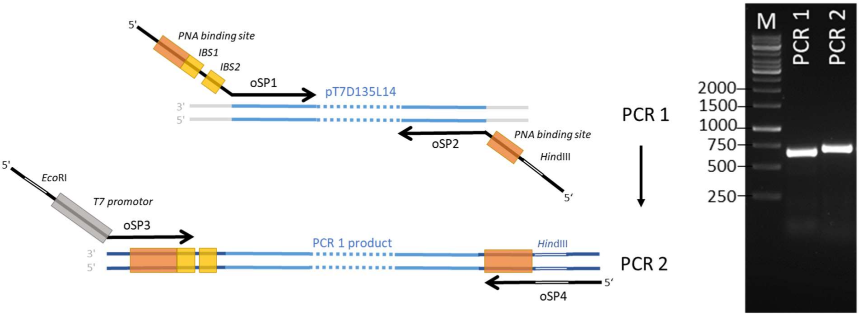


Figure 2.27 Cloning strategy to obtain eD135e-L14 using pT7D135L14 as a template. The agarose gel is showing the right size of the amplified PCR products of PCR1 (659 bp) and PCR2 (693 bp).

The purified product from the second PCR (0.25 µg) and the plasmid pT7D135L14 (330 ng) was digested with the restriction enzymes *Hin*dIII (10 U, Thermo Fischer Scientific) and *Eco*RI (10 U, Thermo Fischer Scientific) for 3 h at 37 °C. To avoid re-ligation, the plasmid was dephosphorylated. After an incubation time of 2 h, 1 U of alkaline phosphatase (FastAP, Thermo Fischer Scientific) was added to the plasmid. Both digested plasmid (3126 nt) and insert (686 nt) were analyzed and purified via an agarose gel and by the PCR clean-up kit (Wizard® SV Gel and PCR Clean-Up System, Promega). The conditions for the ligation of both fragments were the following: digested and dephosphorylated plasmid (24 ng), digested insert (25 ng), and 2.5 U of T4 ligase (Thermo Fischer Scientific) were incubated for 1 h at 22 °C. A ligation sample containing the plasmid alone was used as a negative control. The ligation mixtures were directly used for heat shock transformation into chemically competent *E. coli* (homemade, Xl1blue). 5 µL of each ligation sample was transferred to 50 µL of Xl1blue. The reaction mixture was chilled for 2 min on ice before, and for 5 min after the heat shock of 1 min at 42 °C was applied. The cells were recovered in 100 µL LB-medium for 45 min at 37 °C. The whole transformation mixture was plated on LB agar plates containing ampicillin (100 mg/ml) and incubated overnight at 37 °C. After successful transformation, a few colonies were chosen for isolation of the plasmid DNA by PureYield^{TM} Plasmid Miniprep System (Promega). The sequencing confirmed the correct sequence of the newly generated plasmid pSPA1 (3806 bp) which contains the insert eD135e-L14.

2.4.3 RNA preparation

For *in vitro* transcription of the RNA constructs D135-L14 and eD135e-L14, the template plasmids had to be linearized. For this large-scale digestion, 200 µg plasmid DNA (pT7D135L14 or pSPA1) and 12 U of *Hin*dIII were incubated for 3 h at 37 °C in the recommended buffer. The linearized plasmids were confirmed on an analytical agarose gel and purified via a Midiprep kit (Quiagen). The RNAs were transcribed under the following optimized conditions: 15 µg linearized pSPA1, 5 mM of each NTP, 25 mM $MgCl_2$ for D135-L14 and 15 mM $MgCl_2$ for eD135e-L14, 0.01 % Triton X-100 and homemade T7-RNA polymerase were incubated for 4 h at 37 °C in transcription buffer (TABLE 3.10). Preliminary transcription trials revealed for eD135e-L14 that no splicing takes place during transcription (FIGURE 2.28). The transcribed RNAs were precipitated by ethanol precipitation and purified via 5 % denaturing PAGE and subsequent electroelution. The purified RNAs were stored in dH_2O at -20 °C (TABLE 2.4).

Table 2.4 Properties of the two group II intron derivatives, D135-L14 and eD135e-L14. Molecular weight and extinction coefficients were calculated by using the online tool http://www.molbiotools.com/dnacalculator.html.

RNA	Length [nt]	Molecular weight [kDa]	Extinction coefficient [$mM^{-1}cm^{-1}$]	GC content [%]
D135-L14	610	195	~7200	29.2
eD135e-L14	650	208	~7670	29.7

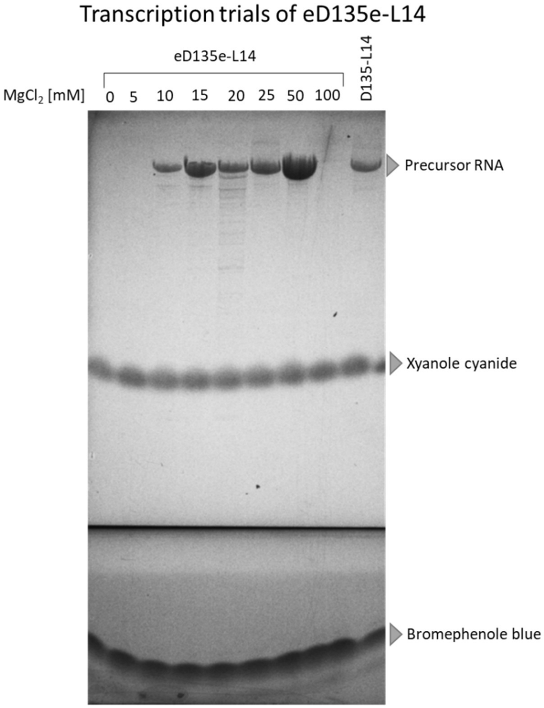


Figure 2.28 8 % denaturing PAGE of transcription trials of eD135e-L14 at different magnesium (II) concentrations. D135-L14 (right) was used as a control. The dyes Xyanole cyanide and Bromophenol blue are indicated. Transcription takes place at a magnesium concentration between 10 and 50 mM MgCl$_2$.

2.4.4 RNA labeling

Both constructs, D135-L14 and eD135e-L14, were labeled fluorescently on the same positions, namely in the artificial loops in domain D1 and D4 (FIGURE 3.57) by using complementary DNA oligonucleotides (TABLE 3.11) or in-house produced PNAs, both carrying fluorescent dyes (TABLE 3.12). D135-L14 was additionally labeled with biotin by hybridizing a complementary DNA-biotin oligonucleotide to the 3'exon. For the activity assays of eD135e-L14, the RNA was labeled on the flanking exons using PNAs.

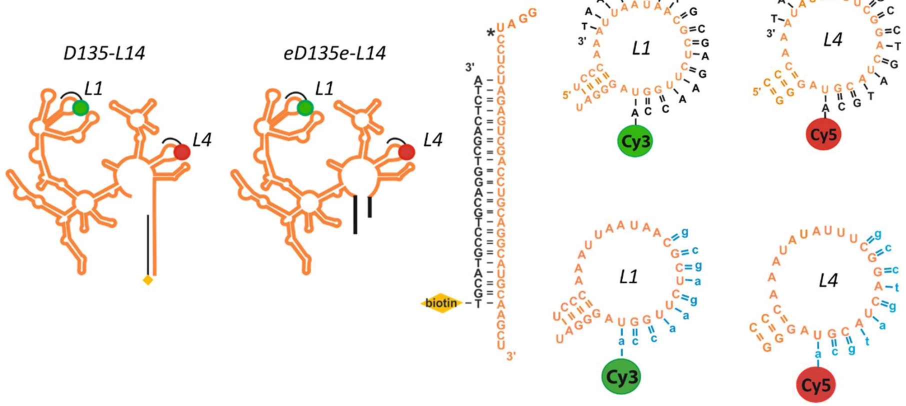


Figure 2.29 Both constructs, D135-L14 (left) and eD135e-L14 (right) derive from the wild-type group II intron Sc.ai5γ. In both cases, the RNA is labeled via artificially introduced loops in domain D1 and D4 (L1, L4). Complementary DNA oligonucleotides (black) or PNAs (blue) carrying a fluorescent dye can hybridize to this labeling platforms. Like this, the RNA is labeled fluorescently. D135-L14 can be directly immobilized on streptavidin-coated surfaces via the biotinylated DNA oligonucleotide complementary to the 3' elongation.

The fluorescent labeling of the RNA was achieved by incubating 1 µM of RNA in a 1:1 ratio with 1 µM of DNA or PNA oligonucleotides each for 1 min at 90 °C in standard buffer (TABLE 3.10). The temperature was decreased to 42 °C for 3 min allowing secondary structures to form and DNAs to hybridize (Schmitz *et al.*, 2015). The folding and/or splicing was induced by adding 100 mM $MgCl_2$. Depending on further proceedings, the folding and/or splicing reaction was continued at 42 °C or stalled at 22 °C.

Table 2.5 DNA oligonucleotide purchased from IBA. Both carry a cyanine dye at their 5'exon. The melting temperatures were determined experimentally by Anita Schmitz (Schmitz et al., 2015).

3'-Sequence	Location	Length [nt]	Tm [°C]
TAA TTA TTG CGA GAA CCA-Cy3	Loop L1	18	57 ± 1
TAT ATA AAG CCT GAT GCA-Cy5	Loop L4	18	56 ± 1
ATC TCA GCT GGA CGT CCG TAC GT-biotin	3'linker	23	67

Table 2.6 PNA sequences synthesized in-house and the corresponding binding location within the intron. The melting temperatures were determined experimentally (Fazliji, 2017).

N to C-Sequence	Location	Length [nt]	Tm [°C]
Cy3-acc aag agc g	Loop L1	10	69 ± 1
Cy5-acg tag tcc g	Loop L4	10	67 ± 1
Cy3-ctg aat gat g	5'exon	10	54 ± 1
gat aca taa t-Cy5	3'exon	10	49 ± 1

2.4.5 Splicing assay on native PAGE

Native gel studies were performed on fluorescently labeled RNA. For the splicing assays, the labeled RNA was incubated at 42 °C for the following time periods: 0, 0.5, 1, 2, 5, 8, 10, 15, 20, 25, 30, 40, 50, 60, 75, 90, 120 and 180 min for magnesium concentrations of 50 mM and 100 mM. For lower magnesium concentrations and for incubation at 25 °C and 30 °C additional time point at 210 and 240 min were taken. The splicing reaction was stopped immediately by adding a final concentration of 50 % glycerol and shock freezing the sample in liquid nitrogen. The samples were loaded onto a native PAGE and run for 2 h at 10 watt at 4 °C. The fluorescent signal of both channels (Cy3, Cy5) was recorded by a typhoon scanner (Typhoon FLA 9500 laser scanner, GE Healthcare). For recording Cy3 signal, the excitation wavelength of 532 nm and a BPG1 filter was used. For recording Cy5 signal, the excitation wavelength of 635 nm and a LPR filter was applied. The photomultipliers were always set to 500 for both channels, and the scan was performed with a pixel size of 50 µm. The data evaluation was performed by using ImageJ. Here, the original tif files were used to get the uncorrected raw data. Every band was selected manually for each population, precursor RNA, intermediate, unbound PNA and cleaved 5'exon (for Cy3 channel) or 3'exon (for Cy5 channel). An overall background was subtracted from each band. The data analysis was performed according to a protocol published by N. Zingler (Zingler, 2014).

2.4.6 RNA encapsulation

A lipid layer was prepared one day prior the experiment by dissolving 1 mg of a lipid mixture (ratio of DMPC to biotin cap PE 99:1, TABLE 3.10) in 100 µL chloroform. The solvent was evaporated under nitrogen for at least 2 h as described elsewhere (Cisse *et al.*, 2007; Liu *et al.*, 2010). The fluorescent-labeled RNA was further diluted to a concentration of 200 nM in a final volume of 100 µL which was used to hydrate the lipids at 30 °C, first for 5 min at 1400 rpm, second for 20 min at 700 rpm. Small vesicles were removed by a centrifugation step for 2 min at 12'000 rpm. The whole mixture was further diluted and well mixed with 500 µL of AB buffer to a final RNA concentration of 33.3 nM. The cloudy suspension was extruded at 30 °C 35 times through a membrane (Avanti Polar Lipids) with a pore size of 100 nm which determined the size of the vesicles. The obtained RNA-vesicle sample was used for immobilization in a microfluidic chamber.

2.4.7 Microfluidic chamber preparation

The microfluidic chambers used for our smFRET studies are homemade. Every chamber consists of two channels allowing two smFRET experiments. For that reason, four holes were drilled inside a quartz slide (7.5 x 2.5 x 0.1 cm, G. Finkenbeiner, Inc.) using a diamond driller (1 mm, Crystalite corporation). After a successively cleaning of the quartz slide and the coverslip (24 x 24 mm, thickness 0.13 – 0.16 mm, Roth) with a plasma cleaner (Zepto One, Diener), both surfaces were aminosilanized according to (Chandradoss *et al.*, 2014), followed by overnight PEGylation (TABLE 3.10). After successive washing with dH_2O both, the quartz slide and the coverslip, were joint together by an imaging spacer (thickness 0.12 mm, SecureSeal^{TM}, Grace Bio-Labs). The chambers were stored in falcon tubes under nitrogen at -20 °C.

2.4.8 smFRET experiment

For conducting smFRET experiments, the RNA was fluorescently labeled and encapsulated into partially biotinylated lipid vesicles as described above (PARAGRAPH 3.4.4). At first, the chamber was flushed once with 200 µL of T50 buffer before coating the surface with 100 µL Streptavidin (20 µL/mL) for 10 min at 22 °C. Unbound Streptavidin was removed by a washing step with 100 µL AB buffer. Either 100 µL of RNA (10 pM, for direct immobilization) or 100 µL of vesicle suspension (for encapsulation) was incubated for 10 min at 22 °C. Meanwhile, IB buffer containing an oxygen scavenger system (OSS) was freshly prepared and used to flush the chamber finally with 200 µL. After an incubation time of 5 min, the chamber was sealed with tabes (3M^{TM}VHB^{TM}) and mounted on a water-immersion objective (NA 1.2m magnification 60x, Olympus), onto the microscope (IX71, Olympus). Immersion oil (type FF, Cargille Laboratories) was used before placing a prism onto the chamber.

The smFRET experiments were carried out on an inverted widefield total internal reflection (TIRF) setup at different temperatures (25 °C and 42 °C) and different exposure times (TABLE 2.7) over 4'000 frames. An area of 512 x 512 pixel, 2 x 2 binning was recorded with the software Andor SOLIS version 4.23. In order to obtain similar fluorescence intensities for both emissions, the power outcome of both lasers (532 nm and 638 nm) was adjusted by changing the optical density (OD) through a neutral density filter wheel (

Table 2.7) prior to every smFRET experiment. Alternating laser excitation (ALEX) with an exposure time of 100 ms, 28 ms, 14 ms or 1.7 ms was applied to every smFRET experiment through acoustic-optic modulators (AOMs) and a controller. The temperature of the objective and the sample holder was controlled by a TC1-100 temperature controller (Bioscience Tools). A bead slide, using fluorescent beads (TetraSpeck™ Microspheres, 0.2 µm, Thermo Fischer Scientific) which are visible in both channels, was recorded once a while and used for adjusting the setup and co-localizing single molecules in both channels.

Table 2.7 Overview of exposure times, laser power, and optical densities. The laser power outcome of both lasers was adjusted by changing the optical density (OD) via the filter wheel according to the individual exposure time to obtain similar fluorescence intensities for both emissions.

Exposure time	Laser 532 nm [mW]	Laser 638 nm [mW]	OD 532 nm/638 nm
100 ms	1.5	0.6	1.9/2.0
28 ms	5.0	1.6	1.3/1.4
14 ms	11.0	4.3	1.0/0.8
1.7 ms	68.0	30.5	0.2/0.0

2.4.9 smFRET setup

For single-molecule FRET measurements, a home-build inverted prism-based TIRF microscope (IX71 from Olympus) was used (FIGURE 2.30) (Zhao and Rueda, 2009). For the excitation path, the two continuous-wave lasers (λ = 532 nm and λ = 638 nm, Crystalaser) were synchronized through acoustic-optic modulators (AOMs, Brimrose) and a controller for alternating laser excitation (ALEX), cleaned-up by bandpass filters (BP, AHF Analysentechnik AG) and the optical density was regulated through different filters within the filter wheel (FW, Laser2000 GmbH). Dielectric mirrors (M) and a dichroic beam splitter (DCM, Laser2000 GmbH) aligned the lasers. The necessary angle of the incident lasers for a total reflection on the quartz surface was achieved through a biconvex lens (L, Newport Spectral Physics GmbH) and the prism (Melles Griot GmbH). The prism is placed on immersion oil (type FF, Cargille) on the quartz slide. The generated evanescent filed of about 200 nm in depth is exciting only surface immobilized fluorophores. An objective and a sample heater (Bioscience tools) were used simultaneously to generate a homogenous temperature over the sample. For the emission path, the emitted photons were captured through an objective (magnification 60x, Olympus) with a high numeric aperture of 1.2 and further spectrally separated by a dichroic beam splitter (DCM) and detected by an electron-multiplying charged-coupled device (EMCCD, Andor iXon). The detected signal was visualized by the software Andor SOLIS version 4.23.

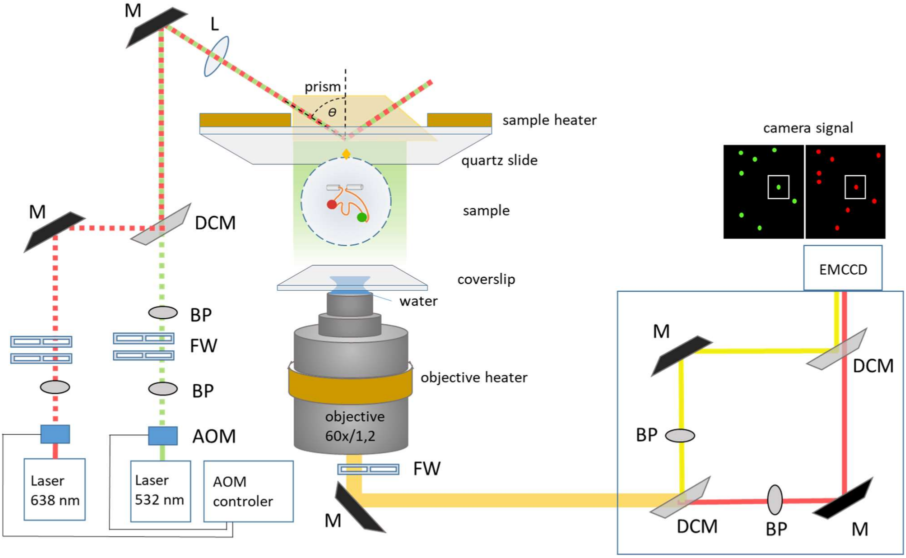


Figure 2.30 Scheme of an inverted prism-based total internal reflection (TIRF) setup for smFRET. Figure adapted from König et al., 2013.

2.4.10 Movie processing and data analysis

Data analysis of smFRET movies was performed by using a multifunctional analysis software for heterogeneous data (MASH), a Matlab base home-written software described elsewhere (Hadzic *et al.*, 2016). For processing a smFRET movie, an average picture of at least 200 frames was generated. On such average image, single molecules in both channels were detected by the spotfinder method houghpeaks and standard parameter. The co-localization of single molecules in both channels was enabled by reference coordinates deriving from a bead slide. After this, fluorescent intensity trajectories were generated in the form of a mash file. Thereby the FRET efficiency and stoichiometry (S) were calculated according to the following equations:

$$FRET = \frac{I_{Cy5}^{532\,nm}}{I_{Cy5}^{532\,nm} + I_{Cy3}^{532\,nm}}$$

Equation 2.4

$$S = \frac{I_{tot,532\,nm}}{I_{tot,532\,nm} + I_{tot,638\,nm}} = \frac{I_{Cy3}^{532\,nm} + I_{Cy5}^{532\,nm}}{I_{Cy3}^{532\,nm} + I_{Cy5}^{532\,nm} + I_{Cy5}^{638\,nm}}$$

Equation 2.5

All further corrections and analysis were conducted on the tab about trace processing. Here, three background corrections were applied: the noise of the EMCCD, the bleed-through of the donor emission into the acceptor channel 0.07 (Hadzic, 2017), and the direct excitation of the acceptor by the donor excitation 0.02 (Hadzic, 2017). We adjusted the fluorescence intensities of both dyes through the laser intensities to obtain similar intensity levels. Like this, the calculated stoichiometry of both dyes is around 0.5, depending on the individual gamma factor. Time traces were manually selected according to i) the presence of both dyes, if both dyes are located within one vesicle in a 1:1 ratio the stoichiometry is 0.5, ii) a stable fluorescent signal over the observation time (no defocusing, blinking, or photobleaching), iii) the surrounding, neighbouring molecules do not interfere, and iiii) mono-encapsulated RNA/dye. Although the probability of multi-encapsulation was reduced by applying an optimal ratio of RNA to lipids, it can occur that either multiple dyes or fluorescent labeled RNAs were encapsulated. Multi-or mono-encapsulation can be distinguished by considering the fluorescent intensities, the stoichiometry and single-step photobleaching (FIGURE 2.31).

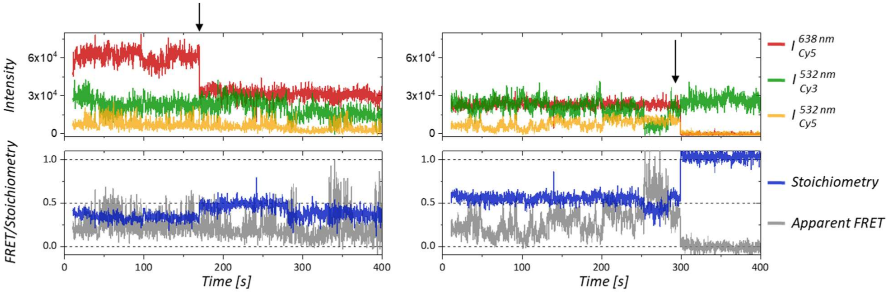


Figure 2.31 Examples of smFRET trajectories showing double- and mono- encapsulation. The fluorescence intensities (red, green and yellow) and the calculated FRET efficiency (grey) and stoichiometry (blue) is depicted. The trajectory on the left shows a bleaching step of Cy5 (red), indicated by the arrow. The fluorescence intensity drops down to half intensity indicating a double-encapsulation of Cy5. The trajectory on the right shows a bleaching step of Cy5 (arrow) as well. Here, clearly, a mono-encapsulation is present, the fluorescence of Cy5 decreases to zero, while at the same time the stoichiometry goes up from 0.5 to 1.

The single molecule trajectories were further categorized into three groups: i) dynamic molecules, which are showing at least one transition between different FRET states. Observing anticorrelation between the donor signal and the FRET signal was an essential requirement. ii) Static molecules are traces without transitions; a static FRET signal above zero is detected over the whole observation time in contrast to iii) the zero FRET traces where both dyes are present, but the FRET value is continuously zero (FIGURE 2.32).

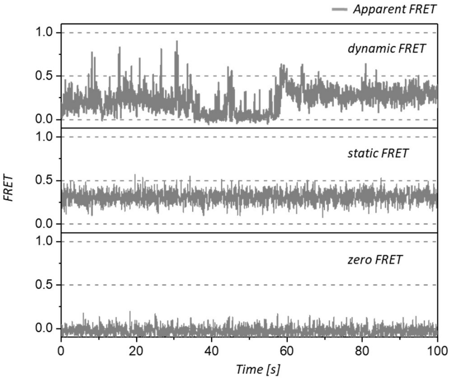


Figure 2.32 Examples of smFRET trajectories of each category. All smFRET trajectories were categorized into three groups: dynamic molecules (top) are showing at least one transition between different FRET states, static molecules (middle) show a constant static FRET over time, and a few molecules show a constant zero FRET (bottom) and a fluorescent signal of both dyes.

All histograms were generated by using OriginPro (2017). The number of bins (50) and the bin size (0.028) were defined for the histograms of the FRET efficiency and the stoichiometry. The fluorescent intensity of Cy5 upon direct excitation ($I_{Cy5}^{638\ nm}$) and the sum of Cy3-signal upon green and red excitation were plotted in histograms separately ($I_{Cy}^{532\ nm}$) (FIGURE 2.33). Here, only the number of 50 bins was set, the bin range was given by the maximal detected intensity in the respective channels.

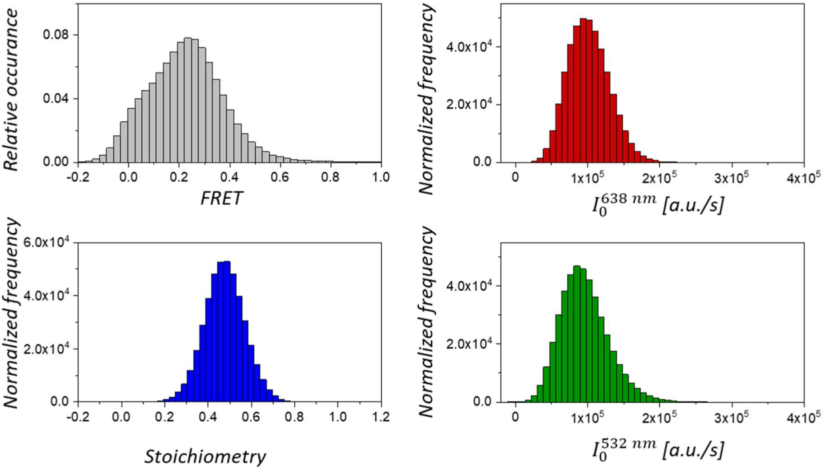


Figure 2.33 Examples of a FRET histogram, histograms of stoichiometry, of the sum of green fluorescence intensities and red fluorescence intensity. The stoichiometry was calculated according to EQUATION 2.5, and the FRET efficiency was calculated by EQUATUTION 2.4.

2.5 Acknowledgements

Special thanks go to Dr. Mélodie Hadzic who initiated the project about RNA encapsulation and did all the preliminary control experiments. Dr. Mélodie Hadzic developed the data analyzing software MASH and continuously optimized it.

Many thanks go to Dr. Richard Börner for all the fruitful discussions, helpful hints regarding smFRET, and correcting this chapter. It was fun to share the enthusiasm for this project.

I would also like to express my gratitude to Dr. Anita Schmitz, Dr. Meng Zhao and Besim Fazliji for synthesizing the desire PNAs.

3. Chapter

Folding and cleavage kinetics of eD135e-L14

Susann Zelger-Paulus, Besim Fazliji, Richard Börner, Roland K.O. Sigel

3.1 Introduction

Non-coding RNA is excised from the coding sequence during protein biosynthesis. In lower eukaryotes, bacteria, plants, and fungi this so-called RNA splicing process is performed by self-cleaving introns (Pyle, 2016). Our focus lies on the visualization of the splicing process of the group IIB intron *Sc*.ai5y, which resides in the mitochondrial genome of *Saccharomyces cerevisiae*. This intron is an ideal example of a perfectly coordinated system where the folding is directly accompanied by splicing. Here, we demonstrate how the covalently attached flanking exons influence the intron folding and cleavage on the single-molecule level and by native gel studies. The comparison between the two intron derivatives D135-L14 lacking the exons and eD135e-L14 including the exons enables to understand the actual underlying process for the first step of splicing and paves the way to investigate the second step, the exon ligation.

Sc.ai5y is a self-cleaving ribozyme located within a housekeeping gene (McNeil *et al.*, 2016). In this context, the precise excision of *Sc*.ai5y followed by the exact ligation of the flanking coding regions during RNA maturation is essential for the host organism. This splicing process underlies a complicated multistep mechanism implying an initial RNA folding step, subsequent intron cleavage, and exon ligation. Although the group IIB intron *Sc*.ai5γ is one of the best-understood introns, a frontier of research is the "lack of information about group II intron dynamics through the stages of splicing" (Pyle, 2016). In order to investigate the individual splicing steps separately, a simplified derivative of Sc.ai5y consisting of three out of six domains was created (Swisher *et al.*, 2001). The truncated derivative D135 comprises all domains essential for conducting the first step of splicing which implies prior RNA folding. To investigate the folding pathway, D135 was further engineered by additional labeling platforms. Therefore, two artificial loops, L1 and L4, were introduced (D135-L14) into the protruding domains D1d2b and D4 (FIGURE 3.35) (Steiner, 2008). Complementary DNA strands carrying a fluorescent dye hybridize to this modifications and thus label the ribozyme fluorescently. The RNA folding of D135-L14 was investigated by single-molecule Förster resonance energy transfer (smFRET) and revealed five folding states; an unfolded, two folding intermediates, a folded state, and a native state (Steiner, 2008; Steiner *et al.*, 2009; Karunatilaka *et al.*, 2010; Fiorini *et al.*, 2015; Fiorini *et al*. in preparation, 2018). During all these studies, the role of the covalently attached exons was always neglected. To investigate their impact on the first step of splicing, we covalently attached short exons to both termini of D135-L14 giving eD135e-L14. Indeed, it could be shown by single-molecule FRET that the exons considerable stabilize the RNA folding (CHAPTER II). In the present chapter, we investigate the influence of the biological relevant exons more in detail. We are expanding the number of labeling positions with the new construct eD135e-L14 from intronic regions (L1, L4) to exonic ones (5'exon, 3'exon) and investigate the interaction between them to create a comprehensive framework of smFRET states (FIGURE 3.34). With the help of distance constraints from the de novo model structure of *Sc*.ai5γ, we disentangle FRET states and assign them to structural motives (FIGURE 3.34-FIGURE 3.37) (Somarowthu *et al.*, 2014).

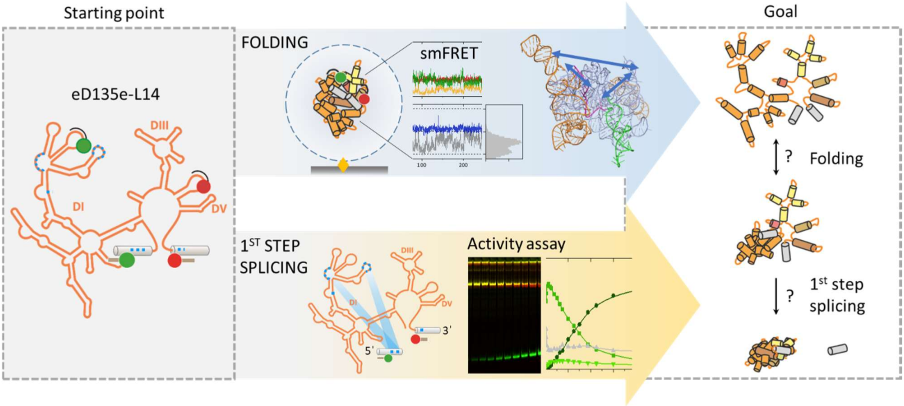


Figure 3.34 Scheme of disentangling RNA folding and the 1st step of splicing of the exon-carrying ribozyme eD135e-L14. The RNA can be fluorescently labeled inside the intron or at its flanking exons. The RNA folding (top) was investigated by applying smFRET on single encapsulated molecules. Then the apparent FRET states are compared with distances of the modeled structure and calculated FRET values (Somarowthu et al., 2014). The cleavage of the 5'exon (bottom) was investigated by cleavage assays. Both processes, folding and cleavage are linked with each other. The presence of misfolded intermediates is still elusive. The goal of this chapter is to understand the impact of the flanking exons on both mechanisms.

Our focus lies on understanding the influence of the flanking exons on folding and cleavage of the group IIB intron (FIGURE 3.34). Although reduced in size, D135 acts still as a ribozyme; it comprises all necessary domains for hydrolytic cleavage. (Su *et al.*, 2005; Swisher *et al.*, 2001; Steiner, 2008; Fiorini *et al.*, 2015; Schmitz *et al.*, 2015). Previously, the first step of splicing was investigated by using the intron derivative D135 and adding its substrate in *trans* (Swisher *et al.*, 2001). This experimental setup obscure two methodology drawbacks. First, the RNA is already pre-folded at the time of substrate addition, so that only substrate cleavage is investigated and second, the single-turnover condition (ratio ribozyme to the substrate is 100:1) underlies a concentration-dependent diffusion of the substrate which reduces the total turnover. Both issues can be overcome by exploring the cleavage activity of the exon-carrying ribozyme eD135e-L14. The cleavage of the covalently bound exons disentangles the preceding folding from the exon excision and reveals main differences of kinetic rates between the *trans*- and *cis*-reaction.

We investigate the folding and exon cleavage of one construct eD135e-L14 applying two different methods. For both, smFRET and activity assays, identical conditions are generated; the RNA is either freely diffusing inside phospholipid vesicles for smFRET or in solution as for native PAGE. This enables to understand the results of the particular experiment, e.g. although the cleavage of the 5'exon is not detectable by smFRET, we know from the activity assays that it has taken place and we can associate changes in FRET histograms with structural changes due to cleavage. To complete the RNA folding studies, we correlate the apparent FRET efficiencies of different label positions with approximate distances, thereby generating a comprehensive picture of RNA folding steps. With the combination of a single-molecule and an ensemble method, we elucidate the influence of the flanking exons on the first step of splicing comprising RNA folding and 5'exon cleavage.

3.2 Results and Discussion

3.2.1 SmFRET folding studies of eD135e-L14

3.2.1.1 *Structure-based FRET model*

FRET is usually referred to as molecular ruler, as the energy transfer efficiency is strongly distance dependent and reports on the distance of the FRET pair introduced in the molecule under study (Schuler *et al.*, 2005). If the molecular structure is known, the distance between two introduced labels, e.g. the FRET pair fluorophores, can be calculated and thus predict the corresponding transfer efficiency (Dimura *et al.*, 2016). The predicted FRET values can further be compared to experimental results. This enables to disentangle FRET-histograms of multi-state systems, such as the here investigated group IIB intron derivative eD135e-L14.

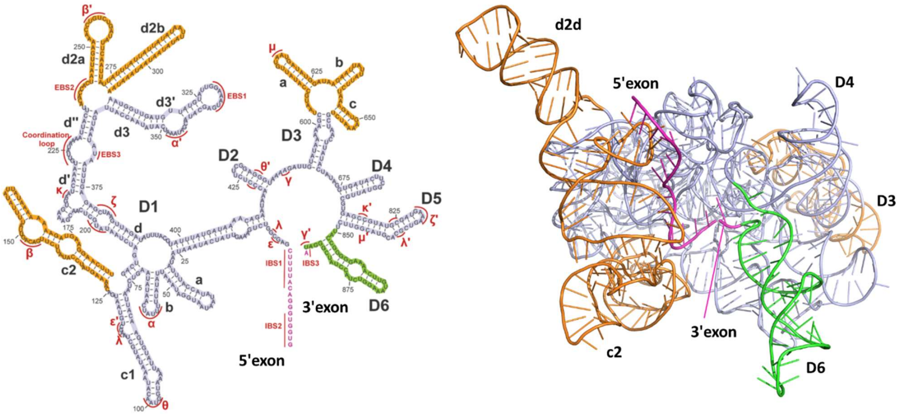


Figure 3.35 Secondary (left) and tertiary (right) structure of the de novo modeled construct eD1356 based on the group IIB intron Sc.ai5γ. The construct eD1356 consists of domain D1, D3, D5, D6, shortened D2 and D4 and is flanked by the 5'exon (magenta) and IBS3 on the 3'exon (magenta). All regions highlighted in grey are similar to OiIIC. Thus, the crystal structure of OiIIC was taken as a template for homology modeling. All subdomains colored in orange and D6 (green) were added to the structure by de novo modeling (Somarowthu et al., 2014). The 3D image was rendered in Pymol using the file nar-02186-r-2013-File010.pdb.

As there is no three-dimensional structure available for the herein used construct eD135e-L14, we further refer to the structure of eD1356, which has been solved via a hybrid approach of homology and *de novo* modeling based on the crystal structure of the homolog RNA sequence of *O. iheyensis* with known crystal structure (Somarowthu *et al.*, 2014). The modeled structure eD1356 comprises original domains D1, D3, D5, D6, shortened domains D2 and D4, as well as a shortened 5'exon of 16 nucleotides and the single nucleotide IBS3 (A888) located on the 3'exon (FIGURE 3.35). The main differences between the here used construct eD135e-L14 and eD1356 are the missing domain D6, the extended flanking exons (5'exon = 31 nucleotides, 3'exon = 14 nucleotides) and the artificial loops L1 and L4 (FIGURE 3.36). These differences are supposed to be not important for the overall fold of the RNA and are thus neglected for the interpretation of fluorophore distances and energy transfer efficiencies.

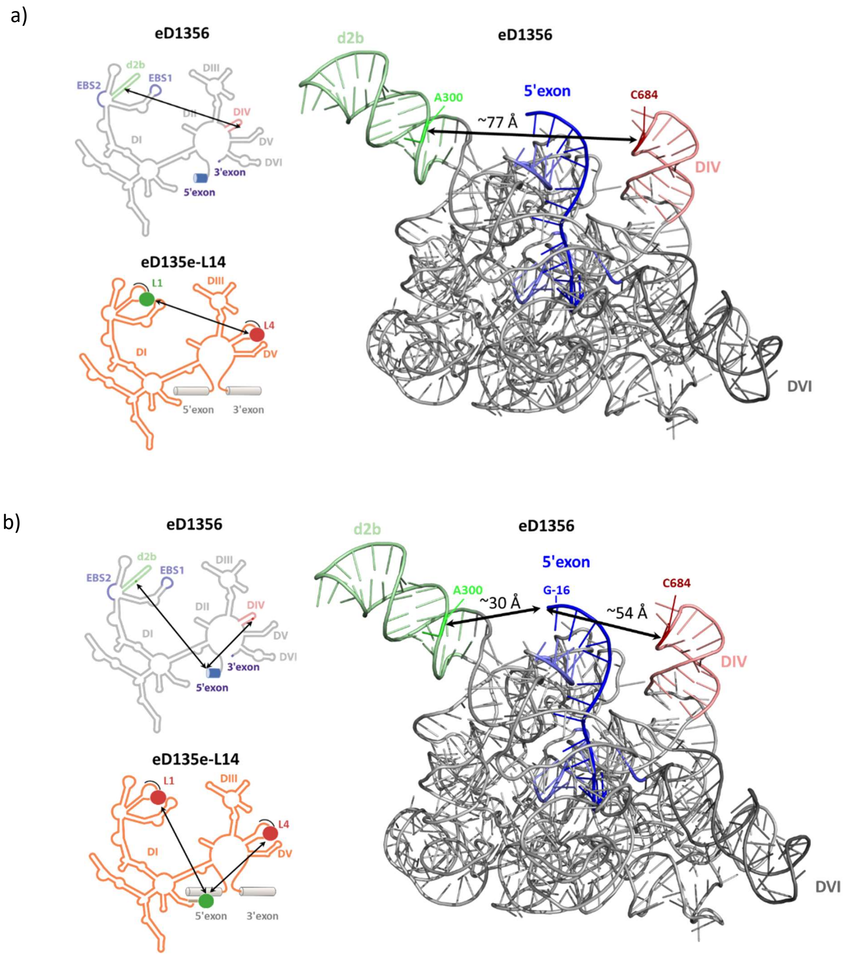


Figure 3.36 Depiction of the FRET labeling scheme for a) loop – loop and b) loop – 5'exon label positions based on the respective secondary (left, upper panel) and tertiary (right) structure of the model eD1356 in comparison with the secondary structure of eD135e-L14 (left, lower panel) (Somarowthu et al., 2014). Domain d2b is colored in light green, domain D4 in light red, both exons are highlighted in dark blue, and the exon-binding sites of the intron are shown in light blue. The remaining domains are shown in grey. ***a)*** *The distance between A300 (bright green) representative for Cy3 and C684 (dark red) representative for Cy5 is approximately 77 Å.* ***b)*** *The distance between A300 (bright green) representative for Cy5 and the 5'exon G-16 (dark blue) representative for Cy3 is approximately 30 Å. The distance between C684 (dark red) representative for Cy5 and the 5'exon G-16 is approximately 54 Å. The 3D images were rendered in Pymol using the file nar-02186-r-2013-File010.pdb.*

In order to determine the transfer efficiency of different label positions in eD135e-L14, we use the particular sequence position of the FRET labels in the corresponding sequence of eD1356. The label position in loops L1 and L4 of eD135L14e correspond to nucleotide A300 in domain d2b and C684 in domain D4 of eD1356, respectively. The 3D model structure of eD1356 shows that both domains, d2b and D4, face out of the RNA structure (FIGURE 3.36A), and are supposed to have no critical structural purpose. The distance between position A300 and C684 is approximately d_{L1L4} =77 Å and corresponds to a theoretical FRET value of $FRET_{\mathrm{L1L4}}$ = 0.11 (FIGURE 3.36A) assuming a Förster radius of 54 Å (TABLE 3.8).

In the second labeling scheme, the introduced 5'exon of eD135e-L14 was labeled fluorescently by PNA-Cy3 without disturbing the important intron-exon binding site interactions. The label position of Cy3 at the 5'exon of eD135e-L14 corresponds to nucleotide G-16 in the homolog RNA. The model structure of eD1356 shows further that the 5'exon is bound to the intron domain D1 via IBS/EBS interactions and is thus located between both label positions L1 and L4 of eD135e-L14 or domains d2b and D4 of eD1356, respectively (FIGURE 3.36B). The distance between the Cy5 label position in L1 corresponding to A300 and the Cy3 label position at the 5'exon corresponding to G-16 was determined to be $d_{\mathrm{L1,5'E}}$ = 30 Å and has a theoretical FRET value of $FRET_{\mathrm{L1,5'E}} \geq 0.9$. The distance between the Cy5 label position in L4 corresponding to nucleotide C684 and the 5'exon G-16 is approximately $d_{\mathrm{L4,5'E}}$ = 54 Å and has a theoretical FRET value of $FRET_{\mathrm{L4,5'E}}$ = 0.5 (TABLE 3.8).

The 3'exon in eD1356 comprised of only IBS3 is positioned very close to the 3'-splice site in the catalytic core formed by domain D1 and D5 (

FIGURE 3.37, right panel). In comparison, the construct eD135e-L14 has an elongated 3'exon, which is 13 nucleotides longer. A second PNA carrying Cy5 is hybridized to that elongation bringing Cy5 in close vicinity to IBS3 at position A888. The distance between the Cy3 label position in L1 corresponding to A300 and the Cy5 labeled 3'exon of eD135e-L14 corresponding to nucleotide A888 is about $d_{\mathrm{L1,3'E}}$ = 75 Å and has a theoretical FRET value of $FRET_{\mathrm{L1,3'E}} \approx 0.12$. The distance between the Cy3 label position of the 5'exon G-16 and A888 is about $d_{\mathrm{5'E,3'E}}$ = 54 Å and has a theoretical FRET value of $FRET_{\mathrm{5'E,3'E}} \approx 0.5$ (TABLE 3.8).

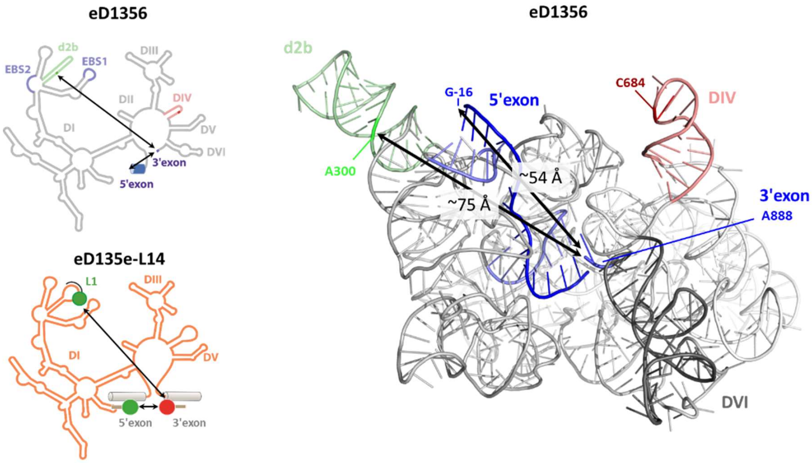


Figure 3.37 Depiction of the FRET labeling scheme for loop L1 - 3'exon and 5'exon -3'exon label positions based on the respective secondary (left, upper panel) and tertiary (right) structure of the model eD1356 in comparison with the secondary structure of eD135e-L14 (left, lower panel) (Somarowthu et al., 2014). Domain d2b is colored in light green, domain DIV in light red, both exons are highlighted in dark blue, and the exon-binding sites of the intron are shown in light blue. The remaining domains are shown in grey. The distance between A300 (bright green) representative for Cy3 and the 3'exon A888 (dark blue) representative for Cy5 is approximately 75 Å. The distance between G-16 representative for Cy3 and the 3'exon A888 is approximately 54 Å. The 3D image was rendered in Pymol using the file nar-02186-r-2013-File010.pdb.

Table 3.8 Summary of fluorescent labels in eD135e-L14 and their corresponding positions in the homology model eD1356, the determined distances, and the corresponding theoretical FRET values.

Label	Position	Distance *d* [Å]	Theoretical FRET
Loop L1/loop L4	A300/C684	77	0.11
Loop L1/5'exon	A300/G-16	30	≥ 0.9
5'exon/L4	G-16/C684	54	0.5
Loop L1/3'exon	A300/A888	75	0.12
5'exon/3'exon	G-16/A888	54	0.5

3.2.1.2 Folding pathway investigated via fluorescence labeling of the internal loops L1 and L4

SmFRET studies of the construct D135-L14 labeled in the loops L1 and L4 comprised five FRET states, an unfolded state *U*, two intermediates I_1 and I_2, the folded state *F* and the most compact state *N* which was attributed to be the native state of the ribozyme (Fiorini *et al.*, in preparation 2018). All previous smFRET studies have been carried out at room temperature (25 °C), and surface immobilized molecules. The native state *N* was shown to be only populated under high salt concentrations (> 40 mM magnesium (II)) and was further stabilized by the addition of the exonic substrate (24 nt long substrate 17/7 in *trans*, comprising the 5'-splice site of intron) or the splicing co-factor Mss116 (Steiner, 2008; Karunatilaka *et al.*, 2010).

In order to study the influence of elevated temperatures and the flanking exons in *cis* on the RNA folding via smFRET, we used the construct eD135e-L14 with different labeling schemes encapsulated in single lipid vesicles. We have shown that the encapsulation approach on smFRET for eD135e-L14 is a suitable technique to follow folding of an activated intron (CHAPTER II). In accordance with previous smFRET studies, we find for eD135e-L14 a stabilization effect of the flanking exons in *cis* as well. In contrast to the literature, the stabilization effect leads to two major populations, namely the two extended intermediate states I_1 at 0.1 and I_2 at 0.25; the *N* state is almost depopulated (CHAPTER II and FIGURE 3.38). An increased population of both intermediates is also visible at elevated temperature 42 °C along with a decrease in the fraction of static molecules (FIGURE 3.38). Further, we find the typical broadening of FRET states due to an increase in RNA conformational dynamics expected for higher temperatures. Former smFRET studies were only conducted at 25 °C, possibly due to the reduced immobilization capacity of the BSA and/or PEG surface passivation scheme for directly surface immobilized molecules. Interestingly, the subsequent decrease in temperature back to the initial 25 °C conditions, completely restores the shape of the FRET histogram as well as the fraction of static and dynamic molecules, supporting the hypothesis of a reversible folding mechanism.

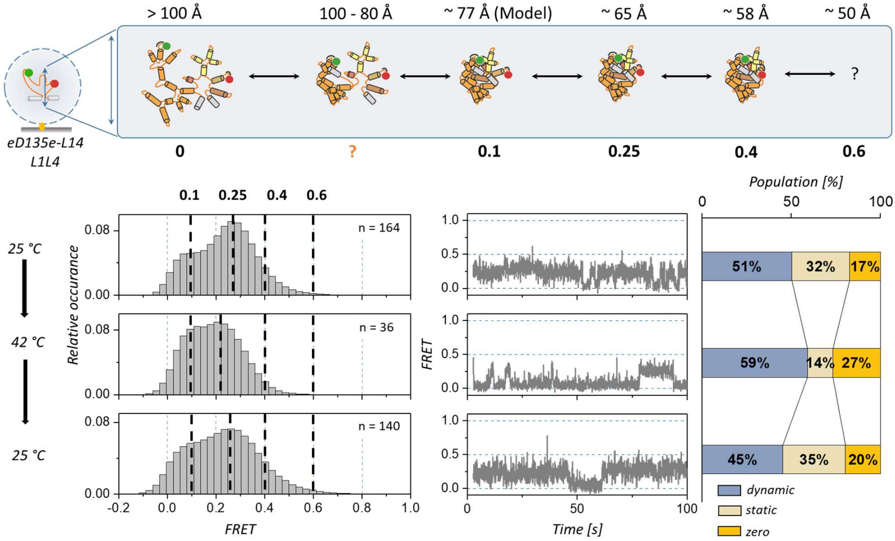


Figure 3.38 smFRET histograms of eD135e-L14 fluorescently labeled with DNA-Cy3 in L1 and DNA-Cy5 in L4 at 25 °C before, at 42 °C during, and back again to 25 °C after ribozyme activation. A correlation between the observed FRET states and the corresponding distances and interpreted folding states is shown on top. A representative trajectory and the amount of dynamic, static and zero FRET states is shown for each condition. Only dynamic and static molecules were taken to build these histograms. The time resolution was 28 ms.

Considering the modeled structure eD1356 and the approximated distance and FRET value, respectively, between L1 and L4 for eD135e-14, we hypothesize that the first extended intermediate state I_1 represents the actual conformation visible in the model structure (Figure 3.38). This is the first time that an experimentally observed FRET value for the group IIB intron ribozyme can be mapped on an underlying three-dimensional structure. Higher FRET states, such as I_2, F and N presumable derive from further compaction of the RNA necessary to perform the catalytic cleavage step or are only due to different loop - loop conformations of the artificially introduces loops L1 and L4.

3.2.1.3 Folding pathway investigated via labeling the flanking exons

In order to visualize the flanking exons during the folding process of eD135e-L14, we labeled them with PNAs. In particular, this labeling scheme allows visualizing if the 5'exon remains bound or dissociates during the ribozyme activation (). At 25 °C, dynamic and static molecules are showing mainly four FRET populations, one broad FRET state between 0.08 and 0.22 which comprises at least two fractions and FRET states at 0.4, 0.65 and 0.85 and higher. The lower FRET states (0.08 and 0.22) are high populated whereas the high FRET states (0.4, 0.65, and 0.85) are low populated. In contrast, after ribozyme activation achieved by temperature increase, the high FRET states are more favored than at 25 °C. A temperature rise leads to higher RNA compaction. The approximate distance from the model structure between the 5'exon bound to EBSI and EBS2 in D1 and the 3'exon pointing into the catalytic triad is $d_{5'E3'E}$ = 54 Å which corresponds to a theoretical FRET value of *FRET* = 0.5 (FIGURE 3.39). This values match with the experimentally determined FRET state at 0.4 and indicates an intermediate conformation, which is favored at 42 °C.

The half of the molecules show a static zero FRET signal. These molecules carry both dyes, but the distance between them is larger than 10 nm. Such a FRET state could refer to unfolded RNAs or to a cleaved 5'exon which is dissociated from the intron. The ribozyme is inactive at 25 °C (confirmed by activity assays FIGURE 3.47). Consequently, the zero FRET can be attributed to an unfolded RNA structure. At 42 °C cleavage has taken place (confirmed by activity assays FIGURE 3.47), now zero FRET can refer to unfolded RNA molecules or RNA molecules where the 5'exon is cleaved and released. The fact that the number of zero FRET molecules is not changing upon ribozyme activation indicates that the 5'exon stays bound after cleavage.

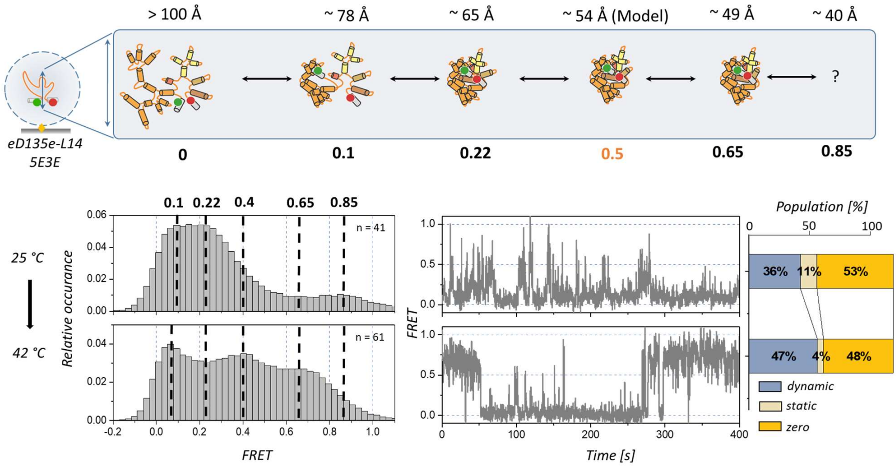


Figure 3.39 smFRET histograms of eD135e-L14 fluorescently labeled with PNAs at both flanking exons at 25 °C and 42 °C. A correlation between the observed FRET states and the corresponding distances and interpreted folding states is shown on top. A representative trajectory and the amount of dynamic, static and zero FRET states is shown for each condition. Only dynamic and static molecules were taken to build these histograms. The time resolution was 100 ms.

smFRET histograms of eD135e-L14 fluorescently labeled with PNAs at both flanking exons at 25 °C and 42 °C. A correlation between the observed FRET states and the corresponding distances and interpreted folding states is shown on top. A representative trajectory and the amount of dynamic, static and zero FRET states is shown for each condition. Only dynamic and static molecules were taken to build these histograms. The time resolution was 100 ms.

Additionally to the evaluation of the FRET states, the single trajectories were sorted into three subclasses, dynamic, static and zero FRET (FIGURE 3.40). The latter was already described above. Dynamic molecules show at least one transition between two FRET states whereas static molecules remain in one FRET state over the observation time (PARAGRAPH 2.4.10). We separated these subclasses to investigate their population and changes during different conditions. The population of zero FRET molecules is dominating the histogram. Excluding these molecules revealed the proportion of dynamic and static molecules. The static molecules are stalled in distinct states for both temperatures. The dynamic molecules occupy the same states as the static molecules do, but the histograms show a broader distribution due to the transitions between the states.

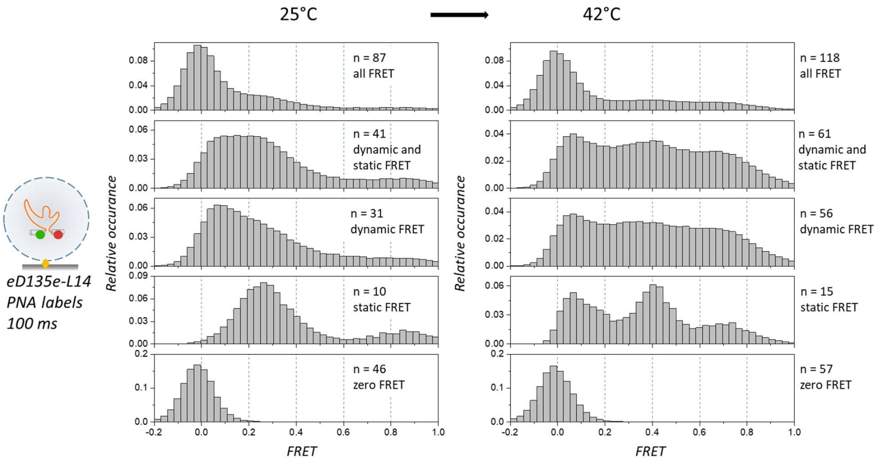


Figure 3.40 smFRET histograms of molecule sorting of eD135e-L14 labeled by PNAs at the flanking exons. The influence of the individual subclasses of dynamic, static and zero FRET molecules on the overall FRET histogram is shown. The time resolution was 100 ms.

3.2.1.4 Folding pathway investigated via cross-labeling the 5'exon and L1

Labeling the 5'exon and the loop L1 in domain d2d give insights into the binding process of the 5'exon to the intron D1 via EBS/IBS interactions. The profile of the FRET histogram in FIGURE 3.41 looks similar to the FRET histograms where the two loops are labeled (FIGURE 3.38), and the 5'exon and L4 are labeled (FIGURE 3.43). The whole histogram looks overall broader and thus more FRET states are visible. The FRET histograms at both temperatures show at least four FRET states, a high populated 0.2, medium populated 0.4 and 0.65 states and a very high FRET at 0.85, which is little populated. The lowest FRET state corresponds to the unfolded RNA where the distance between the 5'exon and L1 is about 68 Å. Both, the 5'exon and L1 get closer in a multistep manner until both bind, and the distance between them shrinks to 30 Å, which refers to a FRET state of 0.97 and corresponds to the homology model.

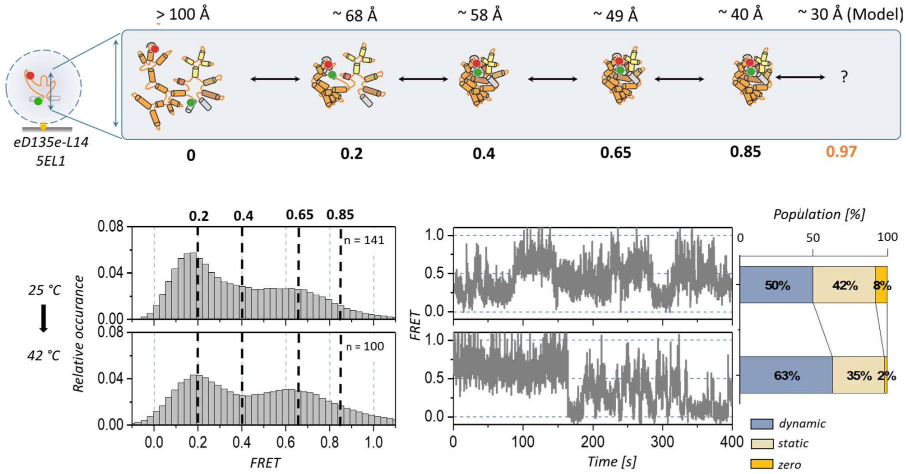


Figure 3.41 smFRET histograms of eD135e-L14 labeled with PNA-Cy3 at the 5'exon and with DNA-Cy5 at L1 at 25 °C and 42 °C. A correlation between the observed FRET states and the corresponding distances and interpreted folding states is shown on top. A representative trajectory and the amount of dynamic, static and zero FRET states is shown for each condition. Only dynamic and static molecules were taken to build these histograms. The exposure time was 100 ms.

Molecule sorting revealed no significant difference between the subclasses and temperature (FIGURE 3.42). Only that, static molecules show three distinct FRET states at 25 °C and only two at 42 °C. The FRET state at 0.4 is becoming apparent at 25 °C whereas at 42 °C the majority of static molecules is stalled in the 0.65 state. The 0.4 FRET state indicates an intermediate between 0.2 and 0.6, which was overcome at a higher temperature. For dynamic molecules, the 0.2 FRET state is the most favored state at both temperatures.

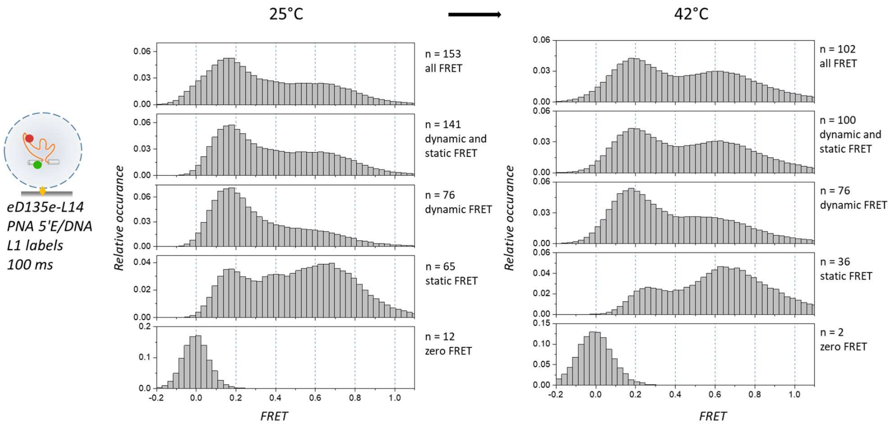


Figure 3.42 smFRET histograms of molecule sorting of eD135e-L14 labeled with PNA-Cy3 at the 5'exon and with DNA-Cy5 at L1. The influence of the individual subclasses of dynamic, static and zero FRET molecules on the overall FRET histogram is shown.

3.2.1.5 Folding pathway investigated by cross-labeling the 5'exon and L4

Previous hydroxyl radical footprinting studies on the folding of *Sc*.ai5γ revealed that the collapse of D1 is the rate-limiting folding step. D1 binds the 5'exon and forms a scaffold to which the other domains are docking. By labeling the 5'exon and L4, we gain insights into this multistep folding process. The shape of the FRET histogram for this labeling scheme (FIGURE 3.43) looks very similar to other labeling schemes, e.g. 5'exon/L1 and L1/L4. Similar FRET histograms showing two main populations for different labeling positions indicates that the overall RNA folding pathway comprises two main structural conformations. The measured distance between the 5'exon and L4 in the structural model is around 54 Å, which corresponds to a FRET value of 0.5. At 42 °C, the 0.6 FRET state shifts to 0.5, which correlates very well to the calculated distance. The experimental data show at least one stable structure, which is less compact than the model structure indicates.

The population of zero FRET molecules is significantly decreasing upon temperature rise they change into dynamic molecules. Zero FRET molecules represent unfolded RNAs at 25 °C, which are transferred into dynamic molecules capable of proceeding folding at 42 °C. A cleavage and dissociation of the 5'exon would lead to the opposite effect, namely an increase in zero FRET molecules. This result is another evidence for a 5'exon, which remains bound after cleavage.

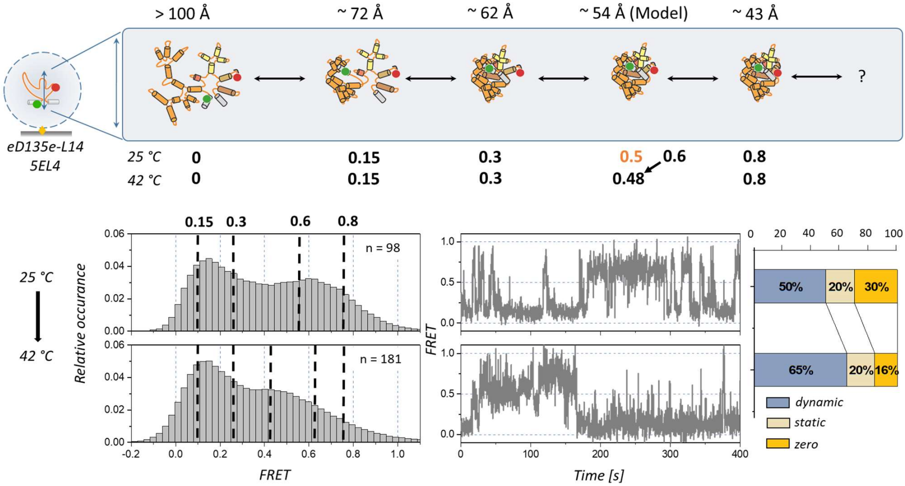


Figure 3.43 smFRET histograms of eD135e-L14 labeled with PNA-Cy3 at the 5'exon and with DNA-Cy5 at L4 at 25 °C and 42 °C. A correlation between the observed FRET states and the corresponding distances and interpreted folding states is shown on top. A representative trajectory and the amount of dynamic, static and zero FRET states is shown for each condition. Only dynamic and static molecules were taken to build these histograms. The exposure time was 100 ms.

Sorting the single trajectories into categories shows that the zero FRET molecules are dominating the overall FRET histogram (FIGURE 3.44). By excluding the zero FRET molecules, we get access to the nature of dynamic and static molecules. Dynamic molecules show an equal distribution of FRET states at both temperatures while the static molecules at 25 °C are mainly stalled in the 0.6 FRET sate.

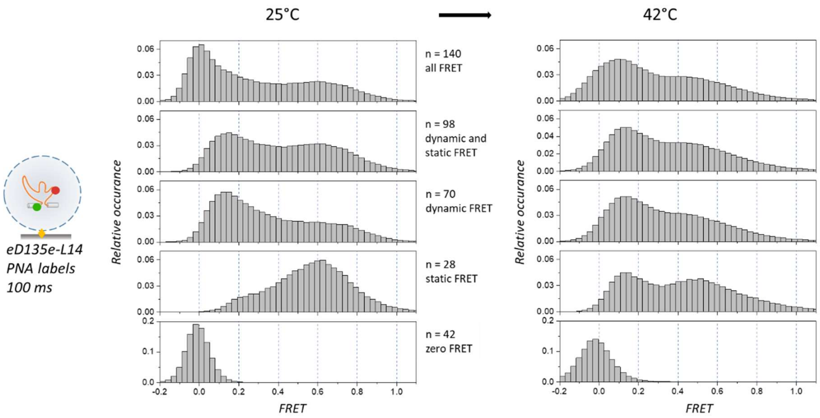


Figure 3.44 smFRET histograms of molecule sorting of eD135e-L14 fluorescently label with PNA-Cy3 at the 5'exon and DNA-Cy5 at L4. The influence of the individual subclasses of dynamic, static and zero FRET molecules on the overall FRET histogram is shown.

3.2.1.6 *Folding pathway investigated by cross-labeling the 3'exon and L1*

Labeling the 3'exon and L1 enables to follow the folding interaction between the 3'exon containing IBS3 and D1 (FIGURE 3.45). If the high amount of zero FRET molecules is excluded from the FRET histogram, the shape of the histogram looks very similar to the one for the labeling scheme of L1 and L4. Two main FRET states are visible at 0.1 and 0.3, and two little-populated FRET states at 0.6 and 0.8. The experimentally determined FRET value at 0.1 corresponds to a distance of around 75 Å, which corresponds to the distance between the 3'exon and d2b in the modeled structure. Higher FRET states might derive from a closer RNA compaction or from local movements of L1.

Ribozyme activation to 42 °C leads to a decrease in the number of zero FRET molecules in favor of dynamic molecules. Again, unfolded RNA molecules turn into dynamic ones, which are capable of proceeding to fold. Even after ribozyme activation by cooling down to 25 °C, the number of zero FRET molecules decrease further. An inactivating of the ribozyme from 42 °C to 25 °C leads to a decrease of zero FRET molecules, which change into static molecules. This result shows that the initially unfolded RNA do not fully unfold after temperature activation, they remain in a partially folded state.

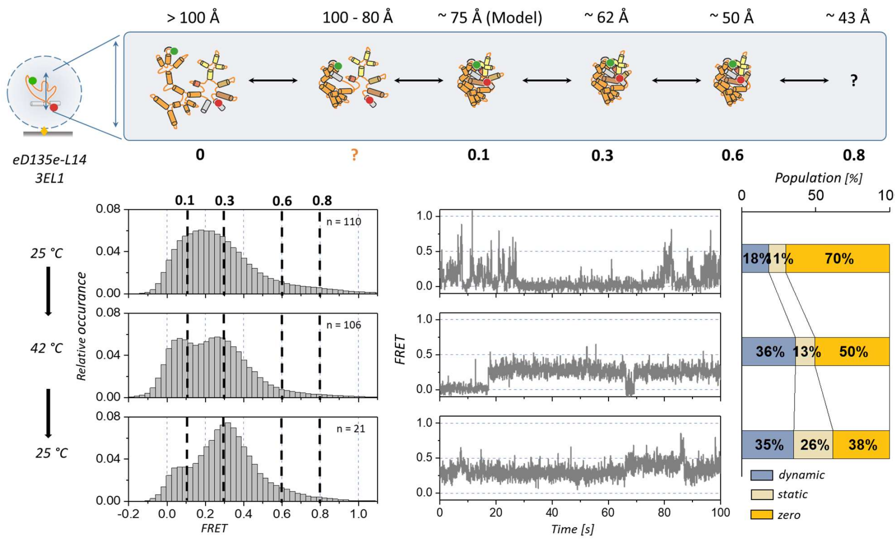


Figure 3.45 smFRET histograms eD135e-L14 labeled with PNA-Cy5 at the 3'exon and with DNA-Cy3 at L1 at different temperatures, before, during and after ribozyme activation at 42 °C. A correlation between the observed FRET states and the corresponding distances and interpreted folding states is shown on top. A representative trajectory and the amount of dynamic, static and zero FRET states is shown for each condition. Only dynamic and static molecules were taken to build these histograms. The exposure time was 28 ms.

Molecule sorting revealed that the FRET histograms are dominated by the zero FRET molecules (FIGURE 3.46). Excluding the zero FRET states shows that the dynamic molecules occupy both main FRET states at 0.1 and 0.3, but after ribozyme activation mainly the 0.3 FRET state is populated. The static molecules are primarily present in the 0.1 FRET state for 25 °C and 42 °C, and after cooling down to 25 °C, both FRET states 0.1 and 0.3 are shown. The 0.3 FRET state is evolving during ribozyme activation for both, dynamic and static molecules. The fact that the shape of the FRET histograms looks different before and after ribozyme activation give evidence of a structural rearrangement, which took place during 5'exon cleavage.

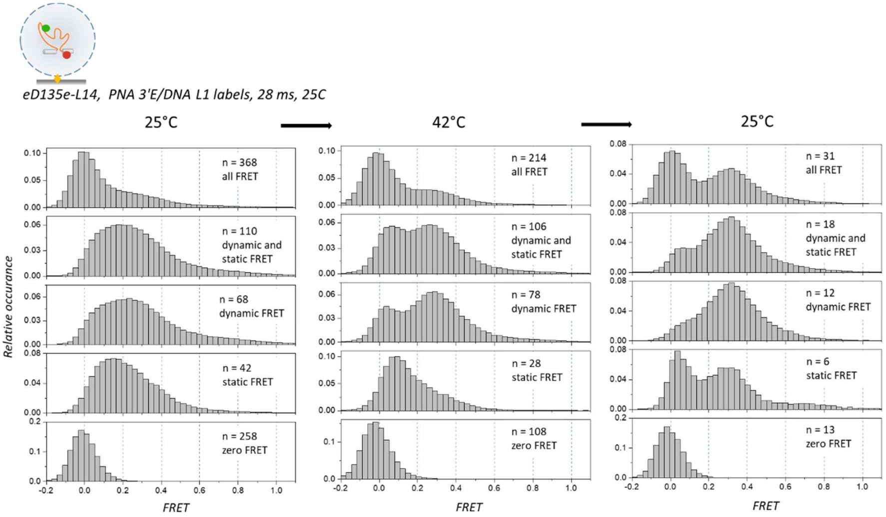


Figure 3.46 smFRET histograms of molecule sorting of eD135e-L14 fluorescently label with PNA-Cy5 at the 3'exon and DNA-Cy3 at L1. The influence of the individual subclasses of dynamic, static and zero FRET molecules on the overall FRET histogram is shown.

3.2.1.7 *Discussion of the folding pathway investigated via different labeling positions*

The modeled structure generated by Somarowthu and co-authors provides a static picture of a group IIB ribozyme in which the RNA is compactly folded (Somarowthu *et al.*, 2014). Domain D1 forms tertiary contacts with other domains and acts as a scaffold for the assembly of them. Therein the 5'exon was found to be bound to D1; domain D3 folds subsequently onto D1 and is finally placed in the vicinity of D5, domain D5 is located in the core of D1 and surrounded by other domains too, and the branch-site adenosine of domain D6 is close to the active site (Somarowthu *et al.*, 2014) (FIGURE 3.35). This structure helps to understand the interplay between long tertiary interactions of the intron at a particular stage of splicing. However, the group IIB intron *Sc*.ai5γ is a dynamic system that adapts different structural conformations to achieve its activity. Which smFRET state belongs to the cleavage-competent state or to an intermediate is still elusive. Thus, the correlation between the experimentally determined FRET states deriving from a dynamic system with the calculated FRET values of the static model is essential and improves the sense of state assignment.

The artificial introduce loops L1 and L4 are located in d2b and D4 (FIGURE 3.35). Both domains are facing out of the complex folded RNA and are supposed to be not directly involved in folding or catalysis (Somarowthu *et al.*, 2014). The first intermediate folding state I_1, which is detected by labeling both loops matches with the calculated distance in the model structure. Thus, we hypothesize that the model structure represents the first intermediate state and higher FRET states like I_2, *F* and *N* derive either from a global RNA compaction or from local domain movements or domain - rearrangements. Such dynamics are not represented by a crystal structure. The fact that the flanking exons stabilize the two intermediate states rather than the native state confirms the hypothesize that the previously assigned *F* and *N* states do not belong to the global folding pathway.

In consideration of the 3D structure, the distance between the fluorophores on the two flanking exons is approximately 54 Å. Both exons are located in the catalytic core. This conformation corresponds to a FRET state which is only favored at 42 °C. At 25 °C, we observe mainly low FRET states where the exons are not in such close proximity. An increase in temperature encourages the position of the exons towards the active site.

Particularly close are the 5'exon and d2b in the model structure, both regions come in vicinity through the formed EBS-IBS interactions. The binding of the 5'exon to D1 is expected to be a two-state binding mode consisting of the preceding folding of D1 and a proper binding of both EBS1 and EBS2 (Su *et al.*, 2005). Our results suggest at least three intermediate states prior to the final conformation. Previous results from (Fedorova and Pyle, 2012) and (Somarowthu *et al.*, 2014) hypothesized already a multistep binding of the 5'exon to the intron. Thereby the more solvent-exposed EBS2 binds first to D1 and initiates a conformational change. Only if EBS2 is correctly bound to D1, EBS1 can bind too. This order of binding might be necessary for specificity and protects from unspecific cleavage. Taking this information, it is very likely that the observed FRET states correspond to a stepwise folding of D1 and simultaneously binding of EBS2 and EBS1. In addition, labeling the flanking exons reveals whether the exons dissociates after their cleavage. If the exons are cleaved and released, the distance between both dyes increase and zero FRET is detected. Here, we observe a decrease of zero FRET molecules upon temperature increase, which indicates that the exons remain bound after excision. The fact that the exons stay bound to the intron during splicing was already assumed after solving crystal structures of an intron at different stages of splicing (Zhao *et al.*, 2015). However, the high effective RNA concentration within the vesicle (3 µM) is another reason why the exons are not dissociating from the intron.

In the model structure, the 3'exon is located close to the catalytic core of the ribozyme that is in the vicinity of the 5'exon. Both EBSs are bound to their counterparts. Labeling the 3'exon and D1 give insights about the interaction between EBS3 and IBS3. The model structure represents the first experimentally determined FRET state at 0.1. Our results indicate another stabilized RNA conformation in which both 3'exon and L1 are even closer than the homology model depicts. This conformation can derive from a global RNA compact in which the 3'exon is allocated for the second splicing step, the exon ligation.

Except for the labeling scheme of both exons, the FRET histograms always show two main populations. The FRET range and the gap between the two maxima differs. Advanced data analysis is necessary to investigate whether these two main fractions belong to the same RNA conformation, e.g. Gaussian fitting and determination of the ratio between the maxima.

Molecules that are showing zero FRET although both dyes can be localized in the vesicle can be attributed to an unfolded precursor or depending on the labeling scheme to a cleaved ribozyme. In any case, they are part of the folding process. If both exons are fluorescently labeled, zero FRET molecules refer to a structural arrangement where the 5'exon and the 3'exon are more than 100 Å away from each other. With a proportion of 53 %, the number of zero FRET molecules is relatively high. In contrast to the labeling scheme of 5'exon and L1. Here, even if the intron is unfolded the distance between both dyes is lower than 100 Å. The number of zero FRET molecules for labeling the 5'exon and L4 is decreasing upon temperature increase. A stable bound 5'exon to domain D1 and thus a decrease in distance to L4 could be the reason. In particular, apparently becomes the high percentage of zero FRET molecules for labeling the 3'exon and L1, which continuously decreases from 70 % to 38 % upon ribozyme activation. The significant change of zero FRET molecules towards dynamic and static molecules demonstrates stabilization of a particular fold where the 3'exon is involved.

3.2.2 Kinetics of the 5'exon of eD135e-L14

eD135e-L14 can undergo the first step of hydrolytic splicing, which includes a preceding RNA folding and a subsequent cleavage of the 5'exon. Here, we investigated the Mg^{2+}-dependent cleavage of the covalently bound exons, made a comparison with the *trans*-reaction of D135-L14, and finally embedded our results into the context of the literature.

The activity of eD135e-L14 depending on the Mg^{2+} concentration over time was determined by labeling the RNA fluorescently. At both flanking exons, fluorophore-carrying PNAs were hybridized. We have chosen the approach of fluorescence labeling to have the same labeling conditions as in smFRET experiments. We can use the gained knowledge from the native PAGE studies for the evaluation of the smFRET experiments regarding, e.g. time- and temperature dependent cleavage.

3.2.2.1 Cis-cleavage assay of eD135e-L14

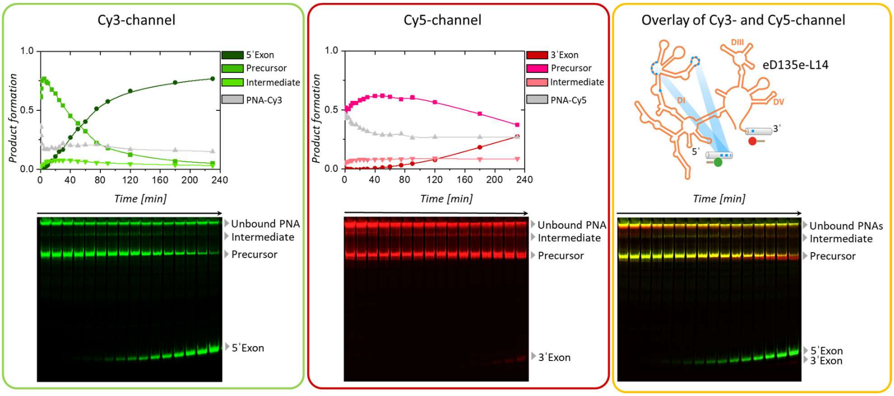


Figure 3.47 Cleavage of the 5'exon of eD135e-L14 over time visualized fluorescently by PNA-Cy3 (left, green) and PNA-Cy5 (middle, red) on a 5 % native PAGE. A co-localization of Cy3 and Cy5 appears in yellow in the overlay of both channels (right). The reaction condition is 0.5 M KCl, 50 mM $MgCl_2$ at pH 7.0 and 42 °C. A native PAGE and the corresponding quantitative data analysis of each detection channel is shown. Four fractions were detected from top to bottom, the unbound PNA, an intermediate, the precursor RNA and the excised 5'exon (Cy3-channel) or the excised 3'exon (Cy5-channel).

We always evaluated the detected emission for both fluorophores. An example of a cleavage assay (500 mM KCl, 50 mM $MgCl_2$, pH 7.0) and the corresponding data evaluation is shown in FIGURE 3.47. In the Cy3-channel, four distinct bands are observed. The band on the top is the fraction of unbound PNA-Cy3 which decreases in the first 20 min from 36 % to 18 % and stays constant afterwards at 19 % ± 2.4. The next fraction appears to be a faint band which corresponds to an intermediate state. The intensity of this intermediate fraction is changing during the incubation time, it increases to a maximum of 7.7 % at 25 min and decreases afterwards again. The next, very dominant band corresponds to the uncleaved precursor RNA which is after an initial increase within the first 5 min, constantly decreasing over time from 77 % to 5 %. While the intensity of the precursor RNA decreases, the fraction of cleaved 5'exon is increasing to 77 % after 230 min of incubation. A further increase upon longer incubation is expected because at time point 230 min a saturation is not yet reached. Also in the Cy5-channel, four main bands are visible. Here, the upper band corresponds to the unbound PNA-Cy5 which decreases continuously from 45 % to 27 %. Followed by the band of the intermediate which increases slowly from 5.7 % to 8.9 % at 10 min and remains unchanged for the rest of incubation. The band of the precursor RNA is split into two, at time point zero a dominant band which corresponds to the full-length eD135e-L14 and a faint band which corresponds to the precursor RNA already lacking the 5'exon (D135e-L14). The differences in size between this two constructs are 31 nt and thus cannot be fully resolved by this native PAGE. Over time, both precursor bands merge into each other. For the evaluation, both bands were considered as one. The intensity of the precursor band is increasing slowly from 49 % to 62 % at 50 min and decreasing afterwards to 37 % at 230 min. The lowest band corresponds to the cleaved 3'exon which starts to increase from 1.9 % at 60 min to 27 % at 230 min. An overlay of both channels shows a co-localization of three bands, namely the unbound PNAs, intermediate and precursor, while the cleaved exons run differently (FIGURE 3.47). To be exact, the overlay of the precursor illustrates nicely a decrease of the full-length precursor (yellow) and an increase of D135e-L14 lacking the 5'exon (red).

3.2.2.2 Correction of unbound PNAs

Although the binding efficiency of PNA is considerably higher than for fluorescently labeled DNA oligonucleotides (Schmitz *et al.*, 2015), still a small fraction of PNA stays unbound. As this unbound PNA fraction is not constant over time, it will be corrected (FIGURE 3.48). Due to its missing charge, the unbound PNA does not migrate in the gel and stays in the gel pocket. The unbound PNA is visible as an upper band in the fluorescent gel. The labeling efficiency for PNA-Cy3 is about 81 % and by this almost 10 % higher compared to PNA-Cy5. PNA-Cy3 has a higher binding efficiency because of two reasons: better accessibility of the 5'exon and a higher melting temperature. Before the data about the Mg^{2+}-dependent cleavage can be fitted, they had to be corrected for the unbound PNA fraction. The correction for PNA-Cy3 is shown in FIGURE 3.48 as an example. The fraction of unbound PNA varies over time. More than 60 % of the PNA-Cy3 is bound to the RNA at time point zero. The amount of unbound PNA follows a characteristic course: it initially rapidly decreases to an absolute minimum, increases again and slowly decreases or stays constant after a maximum is reached. This time course is Mg^{2+}-dependent and is explained by the characteristic folding and cleavage process of the ribozyme. At time point zero, the precursor is the only species present and thus the only location for the PNA to bind. The initial increase of precursor and at the same time the decrease of unbound PNA describes the further binding of the PNA to the precursor RNA. After 10 to 15 min, the RNA is folded into a tertiary complex which makes the exons less accessible for the PNAs. The fraction of unbound PNA stabilizes at around 19 %. After correcting the fraction of unbound PNA for the precursor, the initial fluctuations disappear.

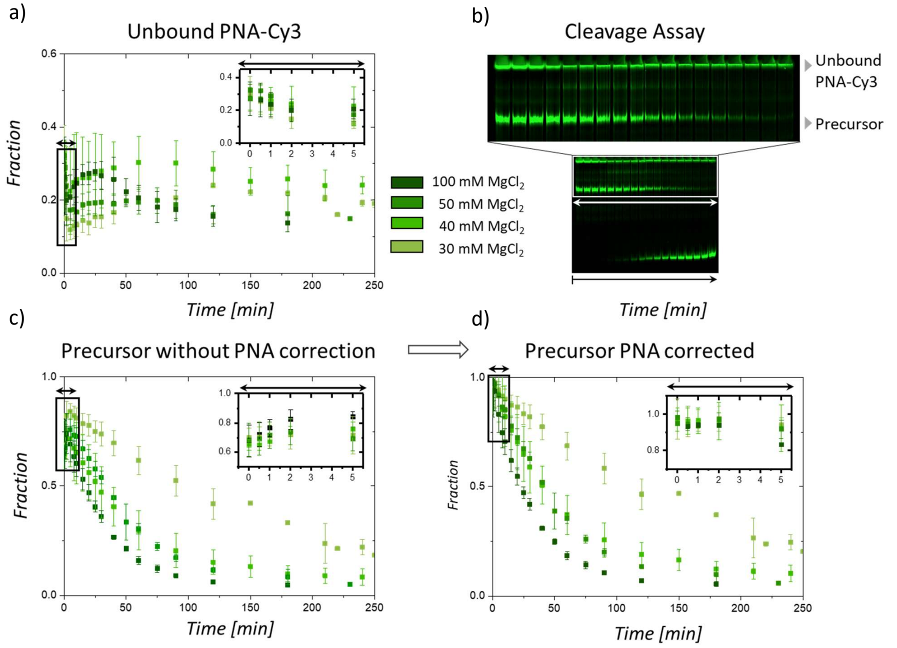


*Figure 3.48 Labeling efficiency of PNA-Cy3 over time at different Mg^{2+} concentrations. The percentage of unbound PNA-Cy3 changes over time, shown in **a)** the evaluated diagram and **b)** the corresponding the native PAGE. An example of a native PAGE at 100 mM MgCl$_2$ is shown. Within the first 15 min, the PNA-Cy3 is still binding to the precursor RNA while at the same time the precursor is excised. For fitting the data, the fraction of the precursor was corrected for the unbound PNA. Diagrams of **c)** precursor before and **d)** after PNA corrections are shown.*

3.2.2.3 Development of an appropriate data fitting model

Different fitting models are applied to the cleavage of eD135e-L14 on the example of 40 mM MgCl$_2$ after PNA correction (FIGURE 3.49). The most straightforward model to fit the data is the single exponential growth model (FIGURE 3.49A) in which the ribozyme-5'exon complex [R5E] irreversible reacts to the cleavage products ribozyme [R] and 5'exon [5E] with *k* as the first order rate constant. This model fits the data of the cleaved 5'exon insufficient, the initial lag phase is completely disregarded, and the intermediate is not included in the fitting model. The series reaction model, which is described by two consecutive exponential functions, fits the data better (FIGURE 3.49B).

Here, prior to cleavage the ribozyme-exon complex [R-5E] undergoes first a structural rearrangement, e.g. from an unfolded to a folded intermediate [I] which is described by the lag phase. This reaction is reversible with $|k_2| = |k_{-2}|$ as the equilibrium rate constants whereas the cleavage with k_1 as the first order rate constant is a one-way reaction. Nevertheless, the intermediate state is poorly fitted. Including an off-pathway to the series reaction which is in equilibrium with the precursor describes the slow continuous increase of the last data points of the cleaved 5'exon and the continued decrease of the intermediate best (FIGURE 3.49C). The off-pathway is attributed to a ribozyme-exon complex [R5E*] which is not capable of cleavage but can reach the cleavage-competent state [I] via [R-5E] with $|k_3| = |k_{-3}|$ as the equilibrium rate constants. In summary, the best fitting model is: the series reaction model including an off-pathway, corrected for the unbound PNA and a global data fitting over all three states, precursor, intermediate and cleaved 5'exon with the shared parameters A_0 (concentration of precursor at time point 0) and I_0 (concentration of intermediate at time point 0) and k_1, k_2 and k_3 attributed to the same process. Further, the folding from [R-5E] to [I] and re-folding from [R5E*] to [R-5E] are in equilibrium and described by k_2 and k_3. The cleavage reaction from [I] to [R] and [5E] is an irreversible step with k_1 as the rate constant. A global fit over all three species and then individually for all Mg^{2+} concentrations fits already excellent. However, global fitting over all Mg^{2+} concentrations and a comparison between both methods would be interesting.

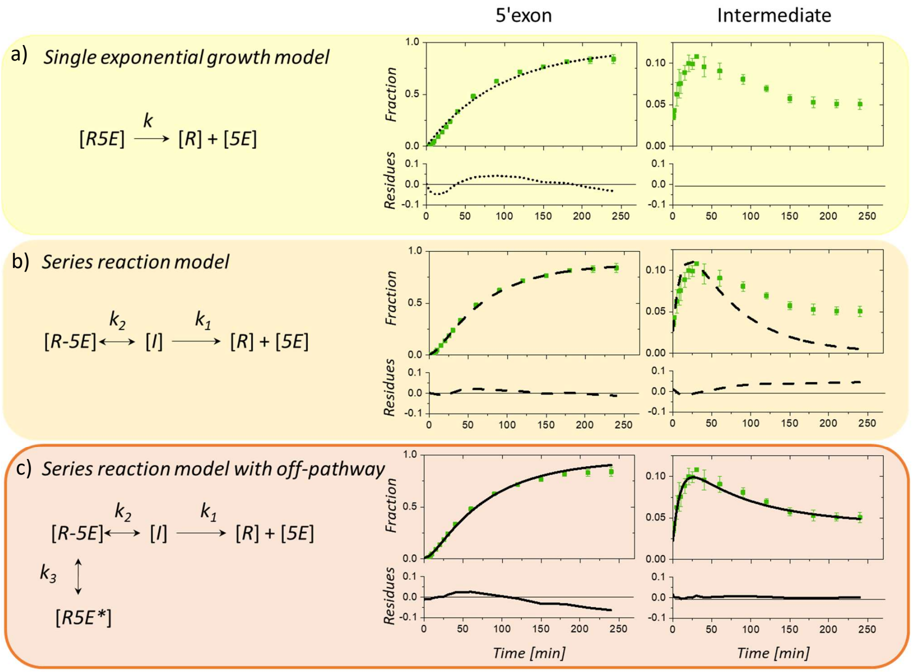


Figure 3.49 Data fitting of the cleavage reaction of eD135e-L14 by the example at 40 mM $MgCl_2$. The fitting of the cleaved 5'exon (left) and the intermediate (right) is shown. ***a)*** *The single exponential growth model is shown in which the ribozyme-exon [R5E] complex reacts to the ribozyme [R] and 5'exon [5E] in one step, described by the rate constant k.* ***b)*** *The series reaction model is describing two consecutive reactions. The unfolded ribozyme-exon complex [R-5E] is folded into a cleavage-competent state [I]. The folding is a reversible reaction described by $|k_2|$ and $|k_{-2}|$ as equilibrium rate constants.* ***c)*** *The series reaction model with the implementation of an off-pathway is shown. A misfolded and inactive ribozyme-exon complex [R5E*] reaches the cleavage-competent state slowly [I] via [R-5E] with $|k_3|$ and $|k_{-3}|$ as equilibrium rate constants.*

3.2.2.4 Magnesium (II)-dependent exon cleavage

The first step of splicing performed by eD135e-L14 is strongly magnesium (II) dependent and confirms previous results (Steiner, 2008; Fiorini *et al.*, 2015). The higher the Mg^{2+} concentration, the faster the cleavage reaction (FIGURE 3.50). The series reaction model described above with an off-pathway included matches perfectly the reaction at all different Mg^{2+} concentrations and for all three subspecies. The fact that the intermediate shows magnesium (II)-dependency as well is another confirmation of playing an essential role in the cleavage reaction.

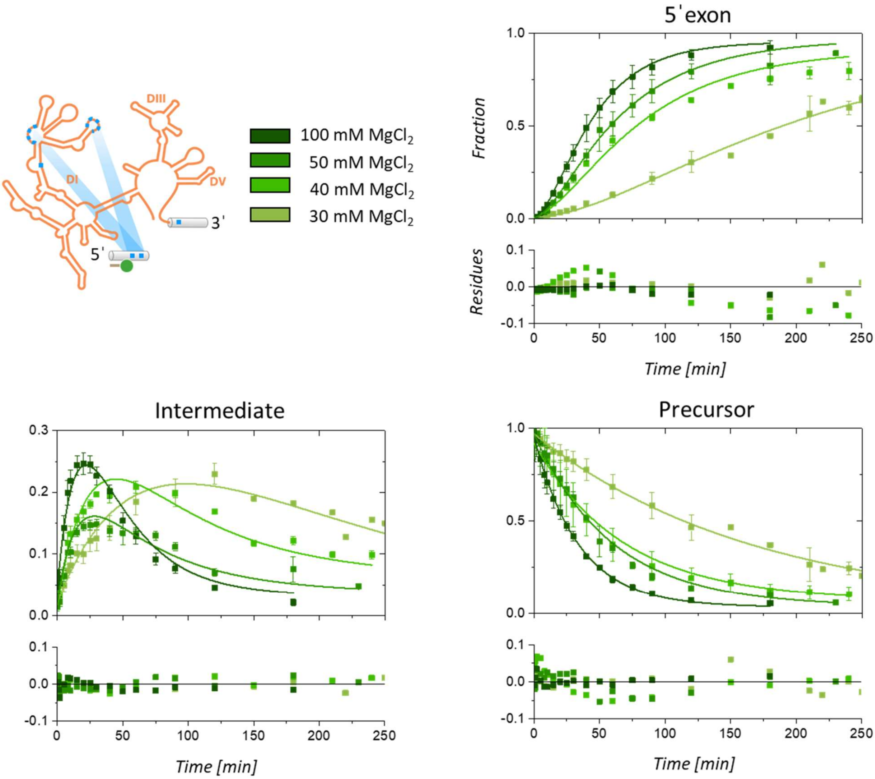


Figure 3.50 Magnesium (II)-dependent cleavage of the 5'exon of eD135e-L14 fitted by the series reaction model including off-pathway (global fitting, precursor, intermediate and off-path in equilibrium). Shown is from left to right the Mg^{2+} dependency of the cleaved 5'exon, the cleaved precursor RNA and the intermediate. The experimental error was determined by triplicate measurements. The differences between the experimental data and the fitted curve are shown as residues underneath every diagram.

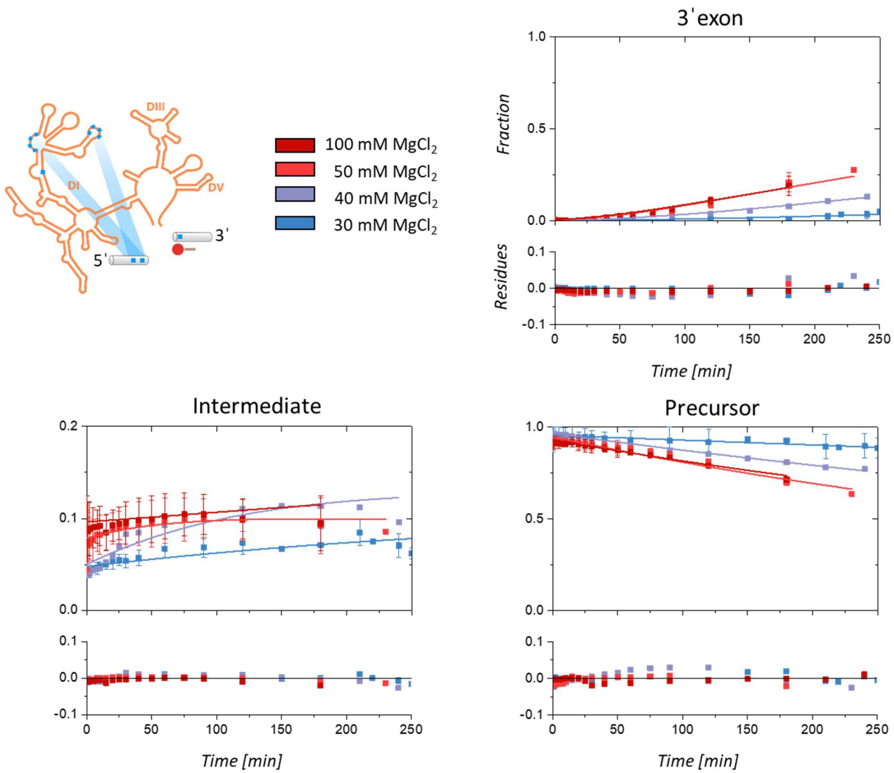


Figure 3.51 Magnesium (II)-dependent cleavage of the 3'exon of eD135e-L14 fitted by the series reaction model including off-pathway (global fitting, precursor, intermediate and off-path in equilibrium). Shown is from left to right the Mg^{2+} dependency of the cleaved 3'exon, the cleaved precursor RNA and the intermediate. The experimental error was determined by triplicate measurements. The differences between the experimental data and the fitted curve are shown as residues underneath every diagram.

The same model as for the 5'exon cleavage can be applied to the cleavage of the 3'exon (Figure 3.51). Also here, the cleavage is magnesium (II) dependent, but the reaction is of one order of magnitude slower. The 3'-splicing site seems to become accessible for a nucleophile attack by a water molecule.

By applying the series reaction model including an off-pathway three rate constants are determined for the intramolecular cleavage reaction of eD135e-L14 (Figure 3.52). The rate constant k_1 describes the fast irreversible cleavage reaction, k_2 characterizes the RNA folding to the cleavage competent state and k_3 is the equilibrium constant of the off-pathway towards the precursor (TABLE 3.17). Whereby the off-pathway is the slowest process. The cleavage of the 3'exon is following the same pathway as for the 5'exon.

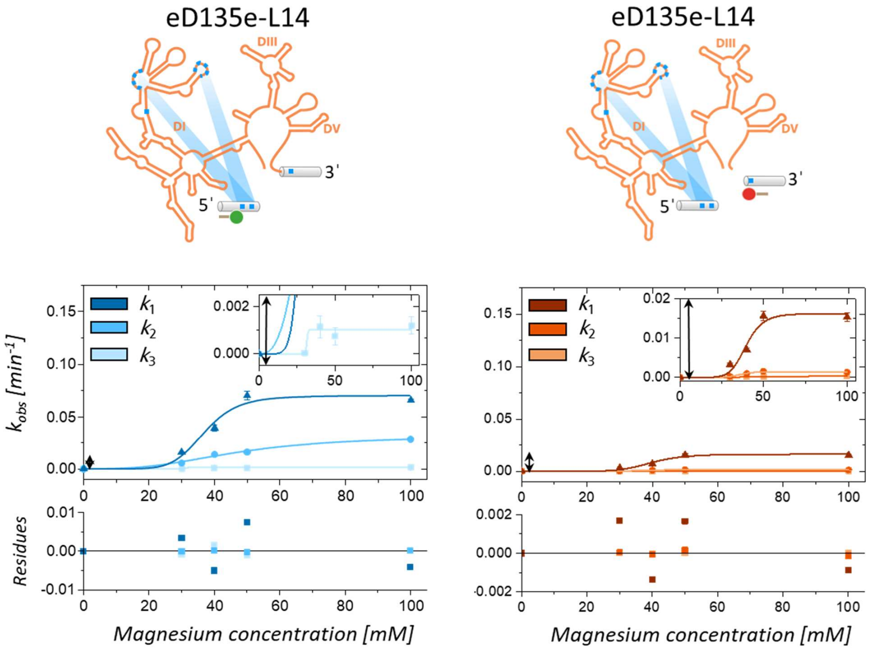


Figure 3.52 Rate constants of the cleavage of the 5'exon (left) and 3'exon (right) of eD135e-L14. The diagram on the left shows three rate constants of the 5'exon cleavage of eD135e-L14 and magnification to visualize the off-pathway. The diagram on the right shows the three rate constants of the cleavage of the 3'exon of eD135e-L14 and a magnification of it. In both diagrams k$_1$ *describes the cleavage reaction, k*$_2$ *the RNA folding towards the cleavage-competent state and k*$_3$ *the off-pathway.*

3.2.2.5 Comparison between trans- and cis-cleavage of the 5'exon

After determining the rate constants of the *cis*-cleavage of the 5'exon, a comparison with the *trans*-cleavage reaction using the literature values is investigated. The activity of D135-L14 was determined by adding the radioactively labeled substrate in *trans* to the ribozyme and follow substrate cleavage over time. It is important to point out that prior the addition of substrate, D135-L14 is pre-incubated for 15 min in the presence of magnesium (II) to ensure proper folding. Substrate cleavage was investigated under single-turnover conditions with a 100-fold excess of ribozyme over the substrate on denaturing PAGE (Fiorini, 2015). The cleavage of the substrate by D135-L14 follows a single exponential decay described by the rate constant k (FIGURE 3.53). In contrast, the activity of eD135e-L14 was determined under native conditions by labeling the exons fluorescently and following the cleavage of the 5'exon over time. Here, the cleavage process was initiated by Mg^{2+}, which reflects time point zero of the reaction. Comparing the two different reactions, intra- *versus* intermolecular process, with each other shows the influence of the covalently bound flanking exons to the reaction (FIGURE 3.53). With regard to the magnesium (II)-dependent cleavage, two main difference becomes apparent between the *trans*- and *cis*-cleavage (FIGURE 3.53, middle). First, the cleavage of the substrate by D135-L14 is by one order of magnitude faster and second, already after 25 min a plateau of about 70 % is reached. In contrast, the reaction of eD135e-L14 takes longer and is almost completed after 120 min by 90 % for 100 mM $MgCl_2$. The rate constant of D135-L14 determined by a single exponential growth model can be compared with k_1 of eD135e-L14 representing the cleavage reaction. Under the assumption that $k_2 = k_3 = 0$, the cleavage reaction described by k_1 follows a single exponential growth model as well and can be compared with the rate constant of D135-L14 (FIGURE 3.53, right). The rate constants were plotted with the following Hill equation:

$$y = y_{max} * \frac{x^n}{K_{Mg}^n + x^n} \qquad \text{Equation 3.6}$$

where y_{max} is the maximum of rate constant k_{max}, K_{Mg} is the Mg^{2+} concentration at half maximum rate and n is the cooperativity constant. The dissociation constant K_{Mg} for the cleavage reaction is similar for both constructs $K_{\mathrm{Mg,\ D135\text{-}L14}} = 37.4 \pm 0.7$ mM and $K_{\mathrm{Mg,\ eD135e\text{-}L14}} = 37.0 \pm 2.3$ mM. For eD135e-L14 an increase of Mg^{2+} beyond 50 mM is not yielding in faster rate constants ($k_{\mathrm{max}} = 0.07 \pm 0.007$ min^{-1}) whereas for D135-L14 $k_{\mathrm{max}} = 0.14 \pm 0.003$ min^{-1} increases by 2-fold.

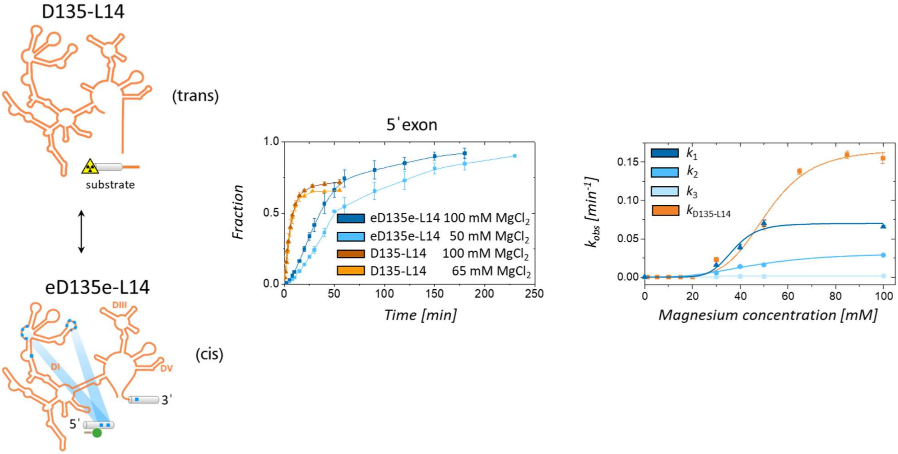


Figure 3.53 Comparison of the cleavage activity between D135-L14 and eD135e-L14. In the diagram in the middle depicts the cleavage of the substrate added in trans to the ribozyme D135-L14 under single-turnover conditions (brown, orange) in comparison with the intramolecular cleavage reaction of the 5'exon by eD135e-L14 (light blue, dark blue). The different rate constants depending on the Mg^{2+} concentration are compared in the diagram showing the hill fit on the right. The excision of the 5'exon in cis is described by three rate constants, k_1 for the cleavage reaction, k_2 for the folding reaction and k_3 for the off-pathway (different shades of blue). The rate constants of the substrate cleavage executed by D135-L14 is shown in orange.

3.2.2.6 Influence of the PNAs on the ribozyme activity

The presence of the two PNAs at the flanking exons could influence the activity of the ribozyme. Thus the RNA was radioactively labeled at the 5'end to investigate the impact of the PNAs, and identical cleavage assays were performed (FIGURE 3.54A). The cleavage of the 5'exon was followed on a denaturing and native PAGE to detect different possible conformations of a fraction. Both gels revealed a precursor band which was decreasing over time and a band of the 5'exon which is increasing simultaneously. No intermediate was detected, all other bands did not change over time and thus could be attributed to the background. The data evaluation showed an identical curve progression of the 5'exon for both conditions. Next, the RNA was simultaneously labeled fluorescently by PNAs and radioactively at the 5'exon. By this, RNA which cannot be detected fluorescently was visualized by radioactivity (FIGURE 3.54B). Interestingly, in addition to the precursor and the cleaved 5'exon, the intermediate became visible in both channels. This experiment shows that the band on top of the fluorescent gel corresponds only to unbound PNAs, no RNA aggregates were detected in the radioactive gel at the same position. The data evaluation of both channels revealed small fluctuations, but overall they were consistent.

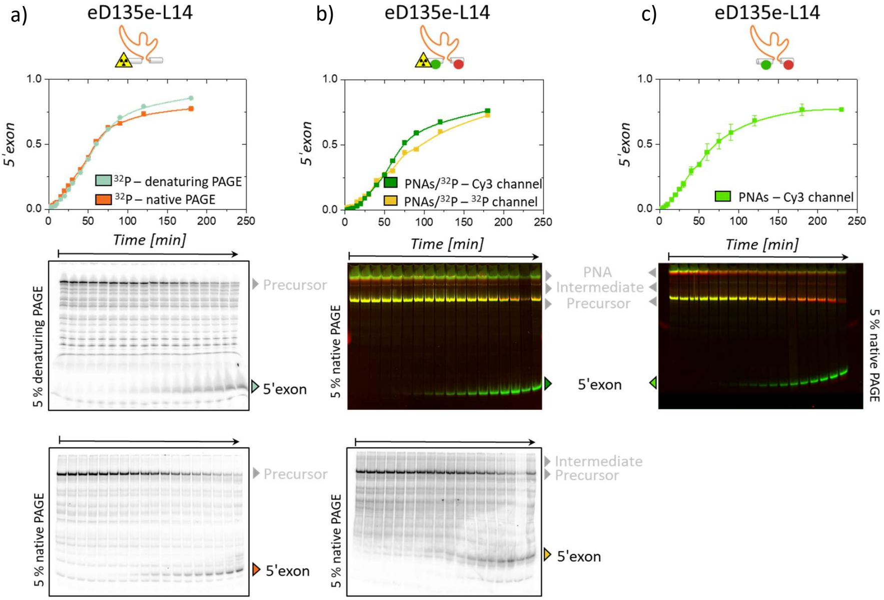


Figure 3.54 The influence of the PNA labels on the cleavage reaction. A comparison between radioactively labeled RNA, PNA- and radioactively labeled RNA and only PNA-labeled RNA is shown. ***a)*** *The cleavage of the 5'exon was followed by labeling the RNA at the 5'end radioactively and following the reaction over time on a denaturing PAGE (top) and native PAGE (bottom). The corresponding data evaluation is shown on top of the gel image.* ***b)*** *The RNA was labeled simultaneously by PNAs (top, overlay of Cy3 and Cy5 channel) and radioactivity (bottom). The data evaluation of both gels is depicted above.* ***c)*** *A representative PAGE and the corresponding data evaluation of a cleavage assay where the RNA was labeled with PNA-Cy3 and PNA-Cy5. The reaction condition was 50 mM $MgCl_2$, 500 mM KCl, pH 6.9 at 42 °C.*

Comparing now the cleavage of the 5'exon between the ribozyme in the absence of the PNAs, in the presence of the PNAs and additionally radioactively labeled, and only fluorescently labeled by the PNAs, the deviations are within the error. All experiments where the RNA was labeled radioactively were carried out only at 50 mM $MgCl_2$. In the absence of the PNAs, the rate constants were determined by applying the series reaction model which are entirely overlapping with the once in the presence of PNAs (FIGURE 3.55).

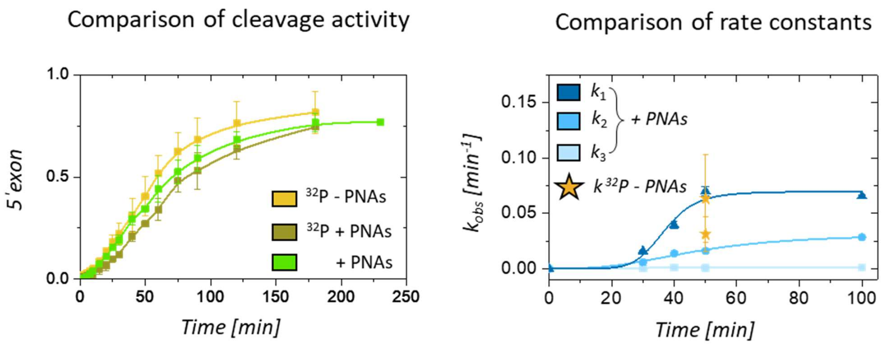


Figure 3.55 A comparison between the cleavage activity and rate constants of eD135e-L14 in the presence and absence of the PNA labels. An overlay of cleavage activity in the absence of PNAs (yellow), labeled with PNAs and radioactivity (brown) and in the presence of PNAs only (green) is shown (left). An overlay of the determined rate constant in the absence of the PNAs at 50 mM $MgCl_2$ (yellow asterisks) with the rate constants in the presence of the PNA labels (blue) is shown in the diagram on the right. The error bars derived from triplicates.

3.2.2.7 Discussion of kinetic results

Kinetic studies on the exon carrying ribozyme eD135e-L14 enable to study the inextricably linked processes of folding and exon cleavage (Figure 3.56). More precise, visualizing the first step of *cis*-cleavage under native conditions by PNA labels revealed four main fractions. After correcting for the unbound PNA labels, three populations per detection channel involved in the cleavage reaction remain. 1) The precursor band is following a magnesium (II)-dependent exponential decay. According to our fit model, the precursor corresponds to a fraction of unfolded RNA, which converts into the cleavage-competent state upon folding. This folding process is reversible and error-prone. Long-range interactions have to be formed correctly. A minority of molecules do not reach the cleavage-competent state straightforward and get trapped in a misfolded state. Refolding takes place from a misfolded stated via the precursor towards the cleavage-competent state. At the same time as the precursor decreases, the amount of cleaved exons is increasing, but not single exponentially. 2) For the 5'exon, a strong magnesium (II)-dependent lag phase is visible before a steep increase of 5'exon takes place followed by a long flat increase. This long and steady increase indicates that the reaction is not completed but slowly continuous. Due to a lack of domain D6, the cleavage of the 5'exon can only be attributed to the hydrolytic pathway in which water acts as a nucleophile. The same is true for the cleavage of the 3'exon. The 3'exon is unexpectedly excised over time as well. Thereby, the cleavage of the 5'exon is one order of magnitude faster than the

3'exon excision. With regard to the splicing reaction, the cleavage of the 5'exon as the first step of splicing is essential for the whole processes, and the cleavage of the 3'exon is just occurring slowly over time. *In vivo*, a separate excision of the 3'exon is very unlikely, because the ligation of both exons is favored. 3) In both channels, the intermediate can be attributed to the cleavage-competent state to which the precursor converts. The cleavage reaction itself is a very fast process; this explains the low occurrence of the intermediate. Once the conformational rearrangement from the precursor to the intermediate has taken place, cleavage occurs immediately. The cleavage reaction can be separated into two reactions, the chemical reaction of the nucleophile attacking the 5'-splice site and the dissociation of the cleaved exon. Disentangling both processes by native gels is impossible. The release of the 5'exon becomes visible on native and denaturing PAGE although it stays bound during splicing, it dissociates from the ribozyme most likely because of applying an electrophoretic field.

The series reaction model including an off-pathway is the model, which fits the data best. The off-pathway is described by an inactive and misfolded state, which is not visible on the gel as a separate band. It has to have different migration properties on a native gel than the precursor RNA. Otherwise, the decay of the precursor would not follow a single exponential function. We assume that the inactive and misfolded state and the intermediate have similar gel migration properties. The inactive state also becomes visible in the slow increase of cleaved 5'exon after the majority of precursor RNA is cleaved. In the absence of an off-pathway, we would expect conversion of precursor and intermediate to the cleaved product yielding in a constant saturation value. Also, the intermediate decays far too slow to be explained by an intermediate only. In the literature, the presence of an off-pathway is still under debate, because it has not been proven (Swisher *et al.*, 2002; Su *et al.*, 2003; Su *et al.*, 2005; Zingler *et al.*, 2010). The fact that such a large RNA adopts incorrect structures during folding cannot be ruled out. Reported lag phases observed in activity assays and remaining inactive fractions are evidence for such misfolded conformations (Daniels *et al.*, 1996; Su *et al.*, 2003; Zingler *et al.*, 2010; Zingler, 2014).

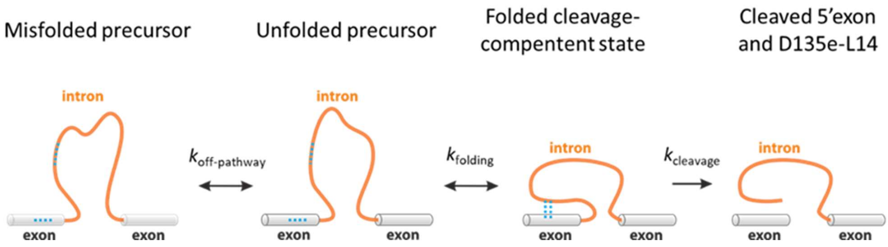


Figure 3.56 A proposed folding and cleavage pathway of eD135e-L14. The majority of unfolded precursor folds straight to a folded cleavage-competent state where the irreversible cleavage of the 5'exon happens. A minority of unfolded precursor is trapped in a misfolded conformation which slowly merges into a cleavage-competent state via the unfolded precursor.

The comparison between the cleavage of D135-L14 of its substrate added in *trans* and the intramolecular cleavage of the 5'exon covalently bound to eD135e-L14 revealed significant differences. The rate constants for the cleavage reaction of eD135e-L14 and D135-L14 are similar to 50 mM Mg^{2+} concentration. At 100 mM $MgCl_2$ the cleavage of the 5'exon of eD135e-L14 is half as fast as for D135-L14 and seems to be limited by the slower process of folding which takes place before cleavage (Fiorini, 2015). A lag phase for D135-L14 is almost not visible because a pre-folding step of 15 min has taken place before the cleavage experiment. Another difference becomes visible in the amount of cleavage product, while for D135-L14 the maximum of cleavage turnover is 65 %, it is about 92 % for eD135e-L14 (FIGURE 3.53). Reasons for that might be the experimental single turnover conditions for the *trans*-reaction, the amount of ribozyme over substrate becomes diffusion dependent, or the percentage of misfolded and inactive molecules is larger for D135-L14 than for eD135e-L14. Similarities of both reactions are the same dissociation constants K_{Mg} of ~37 mM (FIGURE 3.53), and in both cases, a plateau of saturation is not reached.

Although the cleavage reaction of eD135e-L14 is the slowest of all intron derivatives, it is as fast as the wild-type *Sc*.ai5γ covalently linked to its exons (TABLE 3.9) (Zingler *et al.*, 2010). This fact indicates that proper folding is the limiting process in both cases before cleavage can take place.

Table 3.9 The table summarizes the rate constants determined for cleavage and folding, dissociation constants, and the maximum of cleaved product in percentage of different derivatives of Sc.ai5γ. The wild-type construct contains short flanking exons (SE) which differ from eD135e-L14. The data correspond to the following reaction conditions: 500 mM KCl, 100 mM $MgCl_2$, pH 6.9 at 42 °C.

Construct	Cleavage k_{obs} [min^{-1}]	Folding k_{obs} [min^{-1}]	K_{Mg}	% cleavage	Reference
Sc.ai5γ (SE)	0.065 ± 0.015		-	-	(Zingler *et al.*, 2010)
D135	0.41 ± 0.02		-	< 84	(Steiner, 2008)
D135		0.48 ± 0.10	18 ± 1.3[c)] 47 ± 3.7[c)]	< 83	(Swisher *et al.*, 2001)
D135-L14	0.37 ± 0.02		53 ± 3.0	< 84	(Steiner, 2008)
D135-L14 and DNA labels	0.27 ± 0.01		-	< 83	(Steiner, 2008)
D135-L14 and DNA labels	0.16 ± 0.007		37 ± 0.6	< 65	(Fiorini, 2015)
eD135e-L14 and PNA labels	0.07 ± 0.007[a)]	0.03 ± 0.0003[b)]	37 ± 2.0	< 92	This study

[a)] *Rate constant of exon cleavage (k_1)*

[b)] *Rate constant of folding (k_2)*

[c)] *K_{Mg} of folding reaction*

The PNA labels do not influence the activity of eD135e-L14. An overlay of rate constants in the presence and absence of PNAs revealed no differences. Interesting is that only in the presence of the PNAs the intermediate becomes visible. It seems that the PNAs impart different migration properties to the intermediate, which is otherwise hidden underneath the precursor.

3.3 Conclusion and Outlook

Including the flanking exons provides two additional labeling platforms for the same construct. By this, we have now four important reference points to visualize the ribozyme and can follow their interaction with each other by smFRET. Contrary to our expectations, the covalent attachment of the flanking exons to the intron results in the stabilization of both extended intermediate states and not to an increased population of higher FRET states. Thus, we hypothesize that the high FRET states do not belong to the overall global RNA folding pathway nor to a native state. In the context of the modeled structure, which shows a compact folded RNA where the distance between the two protruding labeling platforms L1 and L4 corresponds to the lowest FRET state, high short-lived FRET states can be assigned to local domain movements rather than to states of the global RNA fold. We induce the 5'exon cleavage by a temperature increase and follow structural changes on the single-molecule level. In the vesicle, the 5'exon stays bound to the intron after cleavage, although its cleavage and release are detectable in the splicing assays. The results of the labeling scheme 3'exon/L1 show differences before and after ribozyme activation, which is evidence for a structural rearrangement during cleavage process. The labeling scheme of the 5'exon and L1 revealed the folding of D1 and the subsequent binding of IBS1 and IBS2. The results elucidated at least three intermediates towards the folded domain D1. The fact that the FRET histograms of all labeling schemes show two main populations leads to the conclusion that there are two main RNA folds involved in the folding. Our results have shown that molecules stalled at zero FRET belong to the overall folding path because their population differs when changing the labeling positions and temperature. Including the zero FRET state into the folding path, our results show three different RNA folds: zero FRET which is a mainly unfolded stage, and 0.1 FRET and 0.25 FRET states. Which of both FRET states belong to the active state is still an open question and needs further investigations.

The study on the kinetics of the exon-carrying intron eD135e-L14 reveals three disentangled but closely linked processes; RNA folding, exon-cleavage, and an off-pathway (re-folding step). All three mechanisms are well described by the series reaction model including an off-pathway. The cleavage of the 3'exon underlies the same hydrolytic reaction model as described for the 5'exon, but by one order of magnitude slower. Our results show that the presence of the PNA labels has no significant influence on the ribozyme activity and thus investigating the cleavage fluorescently on native PAGE is a suitable alternative to radioactive labeling. The comparison between the *cis* and the *trans*-reaction reveals that the folding is the rate-limiting step in the cleavage process, and Mg^{2+} can accelerate the cleavage reaction only to a certain extent. A further increase in the rate constant for the *trans*-reaction can be attributed to the experimental conditions. Our study showed that the intron and its exonic substrates function as a unit and investigating them separately in *trans* does not yield the same results.

The presence of the flanking exons enables to study the RNA folding by expanded labeling positions by smFRET and to investigate the kinetics of folding and splicing simultaneously by native PAGE studies. The correlation between the apparent FRET states and an associated structure is of great importance to understand the connection between structure and function. We showed that the high FRET states do not arrive from actual RNA folding, but are induced by the flexibility of domains. The kinetic studies revealed that the folding of the ribozyme is the cleavage-limiting process. Thus, it would be of high interest to investigate the kinetics of the *cis*-splicing and exon ligation of the wild-type *Sc*.ai5γ the same way.

3.4 Materials and Methods

3.4.1 Buffers and kits

Table 3.10 Overview of all buffers and their composition, which were used to perform all experiments of this chapter.

Solution name	Composition	Ref.
Acrylamide gel (denaturing)	7 M Urea, TBE buffer, 8 % Acrylamide, 10 % Ammonium persulfate, TEMED	
Acrylamide gel (native)	5 % Acrylamide, native gel buffer	
Elution buffer	250 mM NaCl, 10 mM MOPS, 1 mM EDTA (pH 6.0)	
TAE buffer	40 mM Tris-base (pH 8.3), 40 mM Acetic acid, 1 mM EDTA	
TBE buffer	89 mM Tris-borate, 2 mM Na_2EDTA (pH 8.3)	
Transcription buffer	40 mM Tris-HCl (pH 7.5), 40 mM DTT, 2 mM Spermidine	(Gallo *et al.*, 2005)
ME buffer	10 mM MOPS, 1 mM EDTA (pH 6.0)	
Native gel buffer	66 mM HEPES, 34 mM Tris (pH 7.4)	
Urea Loading Buffer	12 M Urea, 40 mM Tris-HCl (pH 7.5), 0.1 % Xyanole cyanide , 0.1 % Bromephenol blue, 230 mM sucrose, 0.8 mM EDTA (pH 8.0)	
smFRET experiments		
Anti-blinking buffer (AB)	0.5 M KCl, 80 mM MOPS (pH 6.9), 0.1 M $MgCl_2$, 3 mg/mL Trolox	(Ha, 2001)
Imaging buffer	SAB, 21.7 U catalase, 1.65 U glucose oxidase	(Ha, 2001)

Lipid composition	9.8 mg/mL DMPC, 2.1 mg/mL biotinyl cap PE, chloroforme	(Cisse *et al.*, 2007)
PEG composition	80 mg MPEG-SVA-5000, 8 mg Biotin-PEG-SVA-5000, 0.1 M Sodium carbonate	(Chandradoss *et al.*, 2014)
Standard buffer	0.5 M KCl , 80 mM MOPS (pH 6.9)	(Pyle and Green, 1994)
Sugar anti-blinking buffer (SAB)	AB buffer, 555 mM D-glucose	(Ha, 2001)
T50 buffer	50 mM NaCl , 50 mM TRIS (pH 7.6)	(Zhao and Rueda, 2009)

3.4.2 Molecular cloning

eD135e-L14 contains the original exon sequences on both sides, 31 nt at the 5'end and 14 nt at the 3'end. The cloning of eD135e-L14 is described elsewhere (Chapter II).

3.4.3 RNA preparation

The *in vitro* transcription and purification of the RNA construct eD135e-L14 (650 nt) is described in Chapter II.

3.4.4 RNA labeling

The intron derivative eD135e-L14 was labeled fluorescently at different positions, namely in the artificial loops in domain D1 and D4 (Figure 3.57) by using complementary DNA oligonucleotides (Table 3.11) or in-house produced PNAs which carry a fluorescent dye (Table 3.12). Beside the internal labeling, the RNA was additionally labeled at the flanking exons. Here, only PNAs were used, because of the limited number of nucleotides free for hybridization.

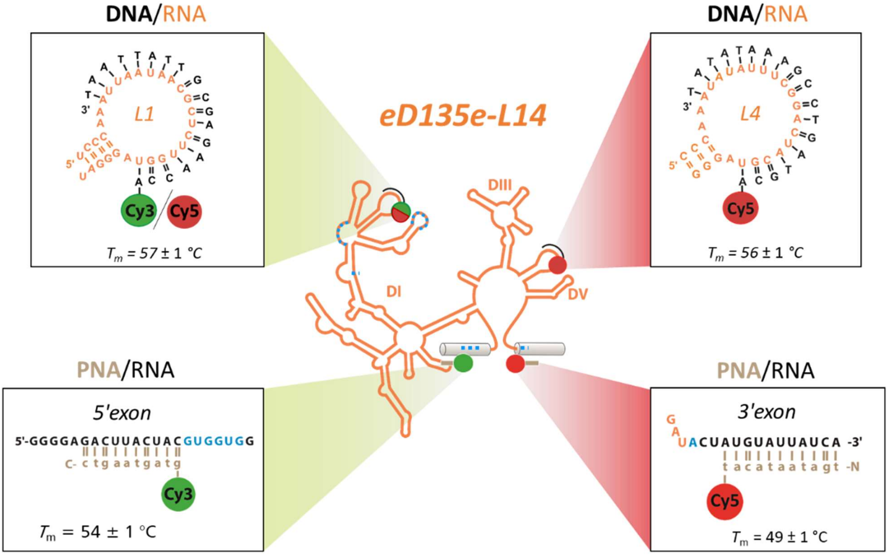


Figure 3.57 The labeling scheme of eD135e-L14. The RNA is labeled via artificially introduced loops in domain D1 and D4 (L1, L4). Complementary DNA oligonucleotides carrying a fluorescent dye can hybridize to this labeling platforms. Like this, the RNA is labeled fluorescently. Additionally, the RNA can be labeled at the flanking exons by PNAs.

The fluorescent labeling of the RNA was achieved by incubating 1 µM of RNA in a 1:1 ratio with 1 µM of DNA or PNA oligonucleotides each for 1 min at 90 °C in standard buffer (TABLE 3.10). The temperature was decreased to 42 °C for 3 min allowing secondary structures to form and DNAs to hybridize (Qin and Pyle, 1999). The folding and/or splicing was induced by adding 100 mM $MgCl_2$. Depending on further proceedings, the folding and/or splicing reaction was continued at 42 °C or stalled at 25 °C.

Table 3.11 DNA oligonucleotide purchased from IBA. Both carry a cyanine dye at their 5'exon. The melting temperatures were determined experimentally by Anita Schmitz (Schmitz et al., 2015).

3'-Sequence	Location	Length [nt]	Tm [°C]
TAA TTA TTG CGA GAA CCA-Cy3/Cy5	Loop L1	18	57 ± 1
TAT ATA AAG CCT GAT GCA-Cy5	Loop L4	18	56 ± 1

Table 3.12 PNA sequences synthesized in-house and the corresponding binding location within the intron. The melting temperatures were determined experimentally (Fazliji, 2017).

N to C-Sequence	Location	Length [nt]	Tm [°C]
Cy3-acc aag agc g	Loop L1	10	69 ± 1
Cy5-acg tag tcc g	Loop L4	10	67 ± 1
Cy3-ctg aat gat g	5'exon	10	54 ± 1
tga taa tac at-Cy5	3'exon	11	49 ± 1

3.4.5 Determination of approximate distances within eD1356

The original sequence of eD1356 was compared with eD135e-L14 containing the artificially introduced labeling platforms L1 and L4 to which the two fluorophores Cy3 and Cy5 are indirectly attached, to determine approximate distances. The nucleotides A300 located in D1d2b and C684 in D4 were chosen to determine the distances among each other in the modeled structure. A300 in eD1356 is positioned similar to U299 in eD135e-L14. Both are 9 nucleotides away from the 4-way junction of D1 (Figure 3.58). U299 is base-paired to an adenine of the hybridized DNA oligonucleotide which carries Cy3. D4 has been truncated to a 16 nucleotide long hairpin comprising a stable tetraloop. C684 is part of the tetraloop and was chosen to mimic the position of the second fluorophore Cy5 (FIGURE 3.58).

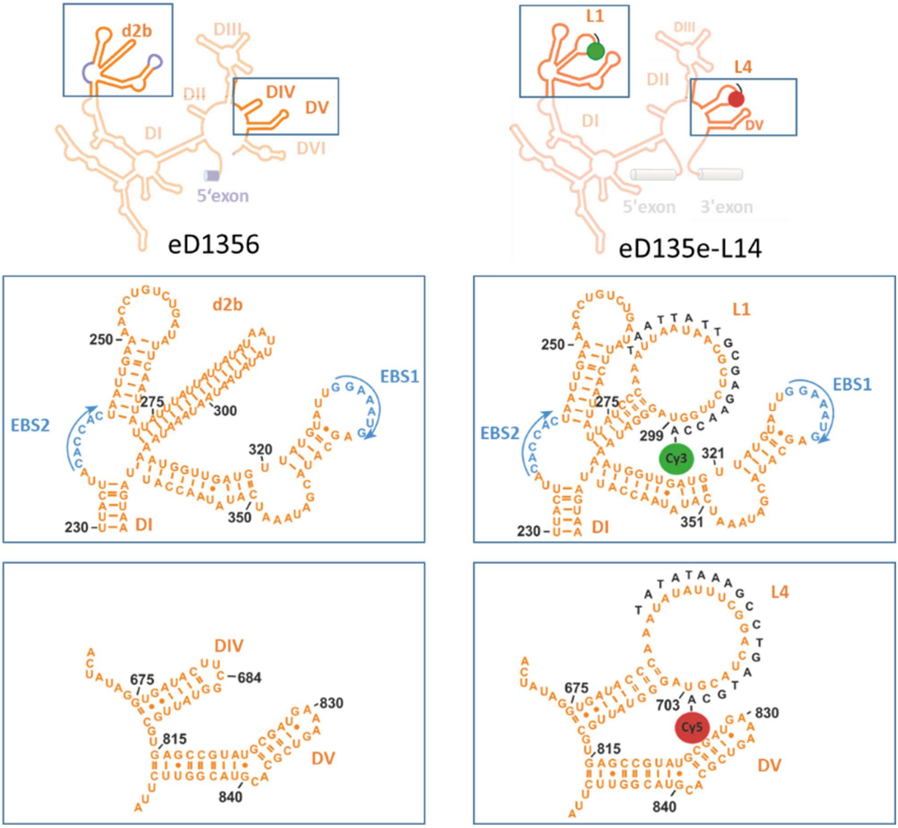


Figure 3.58 A comparison between the original domains of eD1356 (left) with the artificially introduced loops L1 and L4 in eD135e-L14 (right). The nucleotides A300 and C684 in eD1356 are located in the vicinity of the fluorophores in eD135e-L14 and were chosen for an approximate distance determination.

In the modeled structure, the 5'exon is hybridized to the intron via EBS/IBS interactions and thus in very close proximity to A300 of domain D1d2b. The 5'exon is by 14 nucleotides longer in eD135e-L14 compared to eD1356. These 14 nucleotides serve for labeling the RNA by a complementary PNA. The fluorophore Cy3 attached to the PNA is very close to the position of G-16 of the 5'exon. The distance between A300 and G-16 is approximately 30 Å which corresponds to a calculated FRET value of 0.97. The distance between C684 of domain D4 and the 5'exon is about 54 Å which corresponds to a FRET value of 0.5 (FIGURE 3.59).

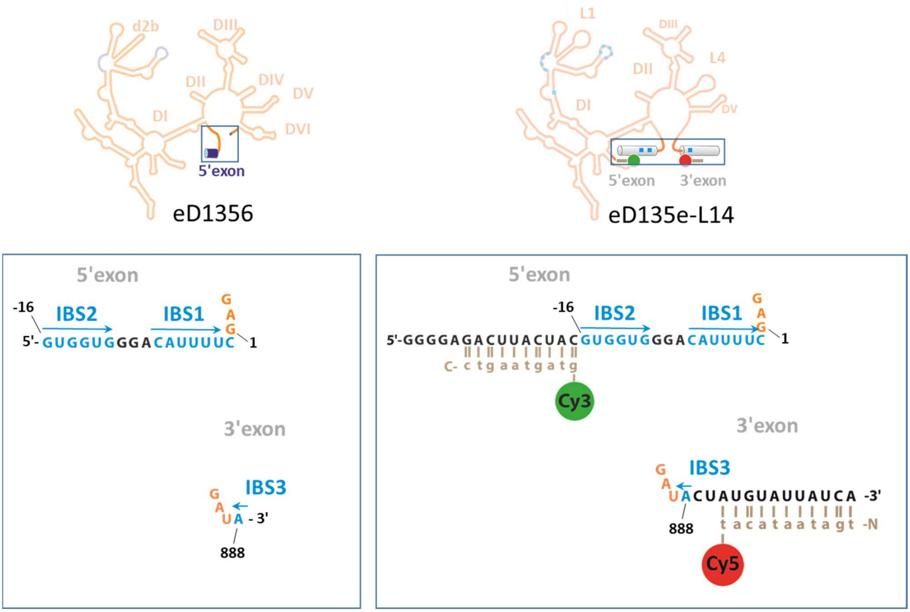


Figure 3.59 A comparison between the exons contained in eD1356 (left) with the 5'exon and the 3'exon in eD135e-L14 (right). The nucleotides G-16 and A888 in eD1356 are located near the fluorophores in eD135e-L14 and were chosen for an approximate distance determination.

3.4.6 ^{32}P 5'-end labeling of RNA

The RNA (eD135e-L14) derives from an *in vitro* transcription and therefore has to be dephosphorylated to add ^{32}P-γ-ATP. 600 pmol of RNA was dephosphorylated in the presence of 4 U of the thermosensitive alkaline phosphatase (Promega) and the appropriate buffer (10x TSAP reaction buffer, Promega) for 1 h at 37 °C. The reaction was stopped by performing a phenol/chloroform (chloroform/isoamyl, 29:1, Sigma) extraction followed by ethanol precipitation. Finally, the RNA was resolved in 60 µL elution buffer, and the concentration was determined by UV-spectroscopy (~ 300 pmol RNA). For labeling the RNA radioactively, 240 pmol of RNA was incubated with 40 U polynucleotide kinase (Promega) and 4 µL of ^{32}P-γ-ATP (6000 Ci/mmol, Perkin-Elmer) in the appropriate PNK-buffer (Promega) for 45 min at 37 °C and 300 rpm. The radioactive labeled RNA was purified by an 8 % denaturing PAGE. The labeled RNA was visualized radiographically using a phosphorimager (Storage Phosphor Screen from Amersham Biosciences, Typhoon FLA 9500 laser scanner from GE Healthcare) and eluted by crush and soak. After additional ethanol precipitation, the concentration of the radioactive labeled RNA (~ 70 nM) was determined by a scintillation counter (Tri-Carb liquid scintillation analyzer 2200 CA from Packard).

3.4.7 Splicing assay on native PAGE

The splicing assays were performed as described elsewhere (CHAPTER II). The following fitting models were applied to determine the rate constants, with the uncleaved ribozyme *[R]*, the flanking exons as substrate *[S]*, the folded cleavage competent ribozyme-substrate complex *[RS]*, the misfolded off-pathway *[R*]* and the cleaved splicing products *[S1]* and *[S2]*:

i. Single exponential growth model, with k as the first order rate constant in min^{-1} describing the cleavage reaction

$$y = A(1 - e^{-kx})$$

ii. Series reaction fitting model, global fit, (Iterationsalgorithm: Levenberg Marquardt) with k_1 as the first order rate constant in min^{-1} describing the cleavage reaction and k_2 and k_{-2} as equilibrium rate constants of the folding process between the precursor and the cleavage-competent state

$$\begin{aligned}
y = {} & s_1*(\mathrm{Offset} + A_0*\exp(-k_1*x))-(s_1 + s_2-1)*((\exp(-k_1*x - k_2*x)*(-A_0*\exp(k_1*x)*k_1 - \\
& B_0*\exp(k_1*x)*k_1 + A_0*\exp(k_2*x)*k_1 + B_0*\exp(k_1*x)*k_2))/(-k_1 + k_2))+s_2*((\exp(-k_1*x- \\
& k2*x)*((A0-\mathrm{Offset})*\exp(k_1*x)*k_1 + B_0*\exp(k_1*x)*k_1 - (A_0-\mathrm{Offset})*\exp(k_1*x + k_2*x)*k_1 - \\
& B_0*\exp(k_1*x + k_2*x)*k_1 - B_0*\exp(k_1*x)*k_2 - (A_0-\mathrm{Offset})*\exp(k_2*x)*k_2 + (A_0- \\
& \mathrm{Offset})*\exp(k_1*x + k_2*x)*k_2 + B_0*\exp(k_1*x + k_2*x)*k_2))/(-k_1 + k_2))
\end{aligned}$$

iii. Series reaction fitting model, off-pathway, global fit, (Iterationsalgorithm: Levenberg Marquardt) with k_1 as the first order rate constant in min^{-1} describing the cleavage reaction, with k_2 and k_{-2} as the equilibrium rate constants between the precursor and the cleavage-competent fold and with k_3 and k_{-3} as the equilibrium rate constants between the precursor and the off-pathway

$$\begin{aligned}
y = {} & s_1*(exp(-(\tfrac{1}{2}) * (2*k_1 + k_2)*x)/k_{\mathrm{eff1}})*(A_0*k_{\mathrm{eff1}}*cosh((\tfrac{1}{2})*k_{\mathrm{eff1}}*x) + (-A_0*k_2 + \\
& 2*k_1*D_0)*sinh((\tfrac{1}{2})*k_{\mathrm{eff1}}*x)) - (s_1 + s_2 -1)*(\exp(-(\tfrac{1}{2})*k_{\mathrm{eff2}}*x)*(2*A_0*k_1^2*k_2*k_{\mathrm{eff1}} - \\
& 2*A_0*\exp(k_{\mathrm{eff1}}*x)*k_1^2*k_2*k_{\mathrm{eff1}} - A_0*k_1*k_2^2*k_{\mathrm{eff1}} + A_0*\exp(k_{\mathrm{eff1}}*x)*k_1*k_2^2*k_{\mathrm{eff1}} + \\
& A_0*k_2^2*k_{\mathrm{eff1}}*k_3 - A_0*\exp(k_{\mathrm{eff1}}*x)*k_2^2*k_{\mathrm{eff1}}*k_3 + 2*k_1^2*k_2*k_{\mathrm{eff1}}*D_0 - 2*\exp(k_{\mathrm{eff1}}*x)
\end{aligned}$$

$$*k_1{}^2*k_2*k_{eff1}*D_0 + k_1*k_2{}^2*k_{eff1}*D_0 - \exp(k_{eff1}*x)*k_1*k_2{}^2*k_{eff1}*D_0 - 2*k_1*k_2*k_{eff1}*k_3*D_0 +$$

$$2*\exp(k_{eff1}*x)*k_1*k_2*k_{eff1}*k_3*D_0 - (1 + \exp(k_{eff1}*x))*k_2*k_{eff1}{}^2*(A_0*(k_1 - k_3) + k_1*D_0) +$$

$$2*\exp((\frac{1}{2})*(k_{eff2}- 2*k_3)*x)*k_{eff1}{}^2*(A_0*k_2*(k_1 - k_3) + B_0*(k_1*k_2 - (2*k_1 + k_2)*k_3 + k_3{}^2) +$$

$$k_1*k_2*D_0))/(2*{}_{keff1}{}^2*(k_1*k_2 - (2*k_1 + k_2)*k_3 + k_3{}^2)))+s_2*(\exp(-(\frac{1}{2})*k_{eff2}*x)*$$

$$(2*B_0*\exp((\frac{1}{2})*(keff_2 - 2*k_3)*x)*(-1 + \exp(k_3*t))*(k_{eff1}{}^2)*(k_1*k_2 - (2*k_1 + k_2)*k_3 + k_3{}^2) +$$

$$A_0*(-2*\exp((\frac{1}{2})*(k_{eff}2 - 2*k_3)*x)*k_2*k_{eff1}{}^2*(k_1 - k_3) + 2*\exp((\frac{1}{2})*k_{eff2}*x)*k_{eff1}{}^2*(k_1*k_2 -$$

$$(2*k_1 + k_2)*k_3 + k_3{}^2) + \exp(k_{eff1}*x)*k_3*(8*k_1{}^3 + 4*k_1{}^2*(k_{eff1} - k_3) + k_2*(-k_2 + k_{eff1})*k_3 +$$

$$2*k_1*(k_2{}^2 - k_{eff1}*k_3)) + k_3*(8*k_1{}^3 - k_2*(k_2 + k_{eff1})*k_3 - 4*k_1{}^2*(k_{eff1} + k_3) + 2*k_1*(k_2{}^2 +$$

$$k_{eff1}*k_3))) + (-2*\exp((\frac{1}{2})*(k_{eff2} - 2*k_3)*x)*k_1*k_2*k_{eff1}{}^2 + 2*\exp((\frac{1}{2})*k_{eff2}*x)*(4*k_1{}^2 +$$

$$k_2{}^2)*(k_1*k_2 - (2*k_1 + k_2)*k_3 + k_3{}^2) + \exp(k_{eff1}*x)*k_3*(8*k_1{}^3 + k_2*(k_2 + k_{eff1})*(k_2 - k_3) +$$

$$4*k_1{}^2*(k_2 + k_{eff1} - k_3) + 2*k_1*(k_2{}^2 + k_2*k_{eff1} - k_{eff1}*k_3)) + k_3*(8*k_1{}^3 - k_2*(-k_2 + k_{eff1})*(k_2 - k_3)$$

$$- 4*k_1{}^2*(-k_2 + k_{eff1} + k_3) + 2*k_1*(k_2{}^2 - k_2*k_{eff1} + k_{eff1}*k_3)))*D_0))/(2*(4*k_1{}^2 + k_2{}^2)*(k_1*k_2 -$$

$$(2*k_1 + k_2)*k_3 + k_3{}^2))-(s_1*s_2-1)*s_3*((1/k_{eff1})*\exp(-(\frac{1}{2})*(2*k_1 + k_2)*x)*$$

$$(k_{eff1}*D_0*\cosh((\frac{1}{2})*k_{eff1}*x) + (2*A_0*k_1 + k_2*D_0)*\sinh((\frac{1}{2})*k_{eff1}*x)))$$

$$k_{eff1} = \sqrt{(4 * k_1^2 \ + \ k_2^2)};$$

$$k_{eff2} = 2 * k_1 \ + \ k_2 \ + \ k_{eff1}$$

3.4.8 smFRET experiments

The RNA encapsulation, the preparation of the microfluidic chamber, the performance of the smFRET experiments including a description of the smFRET setup as well as a detailed explanation regarding data evaluation can be found in CHAPTER II.

3.5 Acknowledgments

At this point, I want to emphasize the fruitful contribution of BESIM FAZLIJI and DR. RICHARD BÖRNER. BESIM FAZLIJI conducted the smFRET experiments of the cross-labeled RNA in the context of his Master thesis under my supervision. Further, he synthesized the PNA-Cy5 and performed the photophysical characterizations of it. Thank you for your help and all the enthusiasm for this research. DR. RICHARD BÖRNER carrrierd out the fittings of the kinetic data and was always a great help in discussing and interpretating the smFRET data.

Many thanks to DR. ANITA SCHMITZ, DR. MENG ZHAO, and the collaboration with the group of PROF. GILLES GASSER for the PNA-synthesis.

4. Summary

The messenger RNA (mRNA) is a single-stranded copy of the DNA that contains all genetic information including coding and non-coding regions. During protein biosynthesis, the coding regions called exons are further translated into an amino acid chain whereas the non-coding regions named introns are excised. The process of intron excision is called splicing. How exactly such a splicing process works is the focus of the presented thesis.

Sc.ai5γ is a sequence residing in the mitochondrial genome of baker's yeast and belongs to a class of introns that can undergo autocatalytic self-splicing. There are structural and mechanistic similarities between this self-splicing introns and the nuclear spliceosomal introns, which indicate an evolutionary relationship and thus makes investigations on *Sc*.ai5γ highly interesting. The group IIB intron *Sc*.ai5γ constitutes an ideal model system in which folding and activity are inextricably linked. The splicing process comprises RNA folding, subsequent intron cleavage, and precise ligation of the two adjacent exons. Among all introns, *Sc*.ai5γ belongs to the best-understood. However, *Sc*.ai5γ as a large, dynamic system undergoes a complex multistep folding and splicing process that is still elusive.

Intron splicing can be achieved via two different pathways, either the branching or the hydrolytic pathway. For the branching pathway, domain 6 (D6) within the intron is essential, because it initiates the first cleavage reaction. In the hydrolytic pathway, a water molecule takes over this role. In any case, the cleavage products are the ligation of the two exons and an excised intron either in the form of a lariat (branching pathway) or linear (hydrolytic pathway). Both pathways are known to occur *in vivo* as well as *in vitro*. For *in vivo* splicing the presence of co-factors is mandatory, whereas *in vitro* splicing can be induced by a high concentration of salt and elevated temperature.

A fascinating feature of group II introns is that they function on a modular base, which allows the combination of different building blocks. In this context, the first step of hydrolytic splicing, namely the cleavage of the 5'exon was studied by using a truncated construct. This minimal active construct D135 consists of three out of original six domains, namely domains 1, 3 and 5. To study the folding of D135 by single-molecule FRET, the RNA was labeled fluorescently with the FRET pair Cy3/Cy5. Therefore, two artificial loops L1 and L4 were introduced into the sites of d2b and D4; both regions are supposed to be not directly involved in folding or catalysis (FIGURE 4.1). Complementary DNA oligonucleotides hybridize to these modifications and thus label the RNA fluorescently.

smFRET was the method of choice to visualize the folding process of D135-L14 over time because it provides access to a heterogeneous system in which rare intermediates or subpopulations exist. Indeed, smFRET studies on the surface-immobilized D135-L14 revealed five folding states instead of the previously discovered three. The RNA folds from an initially unfolded state *U*, through three folding intermediates (I_1, I_2, and *F*) towards the most compact and thus attributed to the native state *N*. Thereby, the presence of high salt concentrations > 40 mM Mg^{2+}, the exonic substrate or the co-factor Mss116 are stabilizing the *N* state. The collapse of D1 is the first and rate-limiting step in the folding process. D1 forms important tertiary interactions with both exons through the intron-exon-binding sites (IBS/EBS) and thus serves as a scaffold to which the other domains are docking (FIGURE 4.1). So far, the role of flanking exons has been neglected in previous smFRET folding studies.

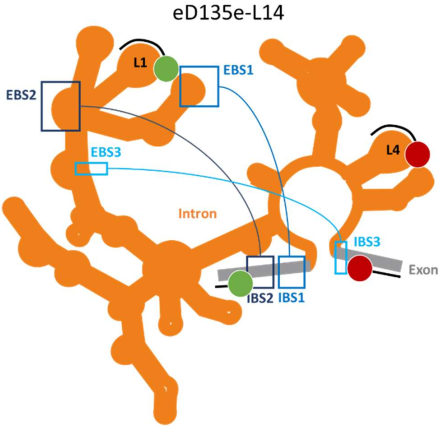


Figure 4.1 Secondary structure of eD135e-L14. The tertiary interactions between the intron and the exon (EBS 1-3 and IBS 1-3) are indicated in shades of blue. In this study, eD135e-L14 is labeled fluorescently at the artificial labeling platforms L1 and L4 by DNA and PNA probes and additionally at the flanking exons by PNAs.

This work is about the influence of the flanking exons on the first step of splicing which implies preceding folding. We developed a new construct eD135e-L14, comprised of the flanking exons, and applied a novel labeling strategy by using fluorophore-carrying PNAs. PNA is a synthetic nucleic acid analog which offers in the context of RNA labeling main advantages over DNA: PNA binds RNA with a higher affinity and specificity. Thus, using PNA labels for smFRET provides a higher labeling efficiency and allows to shorten introduced modifications. Here, we apply PNA and DNA probes to label eD135e-L14 fluorescently within the intron in L1 and L4 and additionally, we use PNAs exclusively for labeling both flanking exons.

4.1 Following RNA folding by smFRET through encapsulation

CHAPTER II focuses on the folding of eD135e-L14, labeled fluorescently with DNAs and PNAs in L1 and L4, investigated by smFRET. Preliminary activity assays of eD135e-L14 revealed a time-dependent cleavage of the 5'exon and additionally a much slower hydrolytic secession of the 3'exon. Because both exons are involved in folding and get cleaved off over time, a surface-immobilization of eD135e-L14 directly via the exons is not possible. An alternative strategy is to trap eD135e-L14 in vesicles which are then anchored on a surface (FIGURE 4.2). Like this, the RNA can freely diffuse inside the vesicle and fold without any steric hindrance. Here, we show for the first time a successful encapsulation of a large ribozyme. By applying the encapsulation technique we were able to investigate the influence of the flanking exons.

Former smFRET studies of D135-L14 observed a stabilization effect of the native state *N* in the present of the exonic substrate. Having the flanking exons covalently attached to the intron shows a stabilization effect of the RNA folding as well, but instead of the *N* state, the folded intermediate states I_1 and I_2 are favored (FIGURE 4.2).

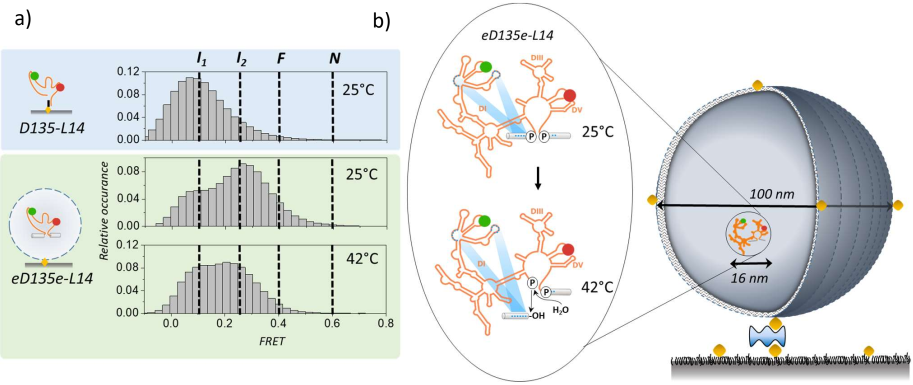


Figure 4.2 smFRET histograms of D135-L14 and eD135e-L14 and the encapsulation scheme. ***a)*** *smFRET histograms of the surface-immobilized construct D135-L14 at 25 °C and the encapsulated construct eD135e-L14 at 25 °C and 42 °C. The FRET states I_1, I_2, F, and N are indicated in every histogram.* ***b)*** *The encapsulation of eD135e-L14 into phospholipid vesicles enables to investigate its folding and ribozyme activation at 42 °C.*

The next step was the activation of the ribozyme. Previously, smFRET studies were always conducted at 25 °C. We showed by activity assays that at such a low temperature almost no cleavage occurs. Thus, we induced the cleavage reaction by increasing the temperature to 42 °C and thereby stimulate ribozyme activity on the microscope on the encapsulated construct eD135e-L14 (FIGURE 4.2). Our results show that the cleaved 5'exon remains bound to the intron and after subsequent decrease of the temperature to 25 °C eD135e-L14 adopts the same folding states as before cleavage. We also observed differences in FRET histograms depending on the type of label, DNA or PNA which can be attributed to the higher flexibility of the loops with the shorter PNA labels.

We showed that encapsulation is a very promising technique for smFRET to investigate a catalytically active RNA, freely diffusing in vesicles. By applying this method, we were able to elucidate the stabilization effect of the flanking exons on the RNA folding.

4.2 Folding and cleavage kinetics of eD135e-L14

The energy transfer between two fluorophores is strongly distance dependent. Therefore, FRET is usually applied as a molecular ruler. The distance between two dyes reveals structural information of the biomolecule they are attached to. Here, we use the homology model of a group IIB intron consisting of a short 5'exon and the domains D1, D3, D5, and D6 to calculate the theoretical FRET value between the labeling positions in eD135e-L14. We found a close match between the theoretical FRET value of eD1356 and the experimental determined intermediate state I_1 of eD135e-L14 with the labels at L1 and L4. This result indicates that the first intermediate state represents the actual conformation of the model structure and higher observed FRET states like I_2, F and N derive from a closer RNA compaction or from loop movements. We show for the first time a mapping of experimentally determined FRET values on a three-dimensional structure.

The presence of the flanking exons provides two additional labeling platforms for eD135e-L14. In CHAPTER III, the folding of the encapsulated construct eD135e-L14 and the influence of the flanking exons were further studied by applying now four different labeling positions and combining them. By this, we mapped for every labeling scheme the observed FRET states to the three-dimensional structure. Our results revealed a dynamic, reversible folding pathway in which the model structure represents an intermediate since we observe closer RNA compaction than the three-dimensional structure depicted. In addition, we included the zero FRET state into the folding pathway, because its population is changing upon different conditions.

In the second part of CHAPTER III we focused on the first step of splicing, namely the cleavage of the 5'exon and determined the kinetics via fluorescent native PAGE. For the kinetic studies, eD135e-L14 was labeled fluorescently at the flanking exons with PNAs. First, a comparison of cleavage rates between the radioactively labeled and PNA-labeled RNA revealed, that the PNAs do not influence the ribozyme activity. Thus, all subsequent activity assays were conducted with PNAs. The kinetic study on eD135e-L14 disentangled three consecutive, interconnected processes; RNA-folding, exon cleavage, and an off-pathway which are well described by the series reaction fit model including an off-pathway. We showed that both, the cleavage of the 5'exon as well as the 3'exon excision underlies the same hydrolytic reaction model, but by one order of magnitude different. Thereby the off-pathway is attributed to a misfolded conformation which reaches a cleavage-competent state over time. The comparison between the *cis* and the *trans* cleavage reaction disclosed divergences, due to different experimental setups (FIGURE 4.3). We showed that both reactions differ in the total cleavage turnover and for rate constants at high Mg^{2+} concentration, but have similar dissociation constants K_{Mg}. A comparison between the rate constants of the wild-type sequence carrying short exons (SE) and all D135-derivatives showed that eD135e-L14 and *Sc*.ai5γ (SE) have similar activity, while the kinetics of e.g. D135-L14 are much faster. This result indicates that folding of the ribozyme is the cleavage-limiting step.

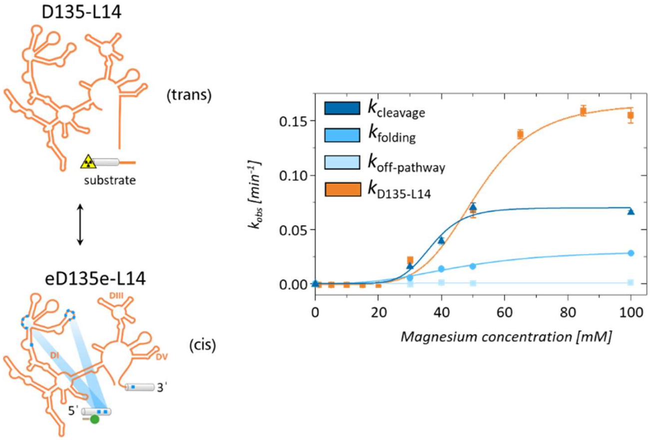


Figure 4.3 Comparison between D135-L14 and its substrate added in trans with the exon-carrying intron eD135e-L14. The three rate constants $k_{cleavage}$, $k_{folding}$, and $k_{off\text{-}pathway}$ describe the reaction of the 5'exon cleavage, because the data were fitted by the series reaction model including off-pathway. In comparison, the substrate cleavage of D135-L14 was fitted with a single exponential model and yield in one rate constant.

5. Zusammenfassung

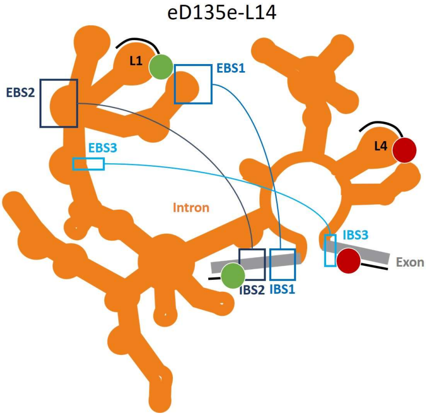
Die Boten-RNA oder auch mRNA genannt (vom englischen messengerRNA) ist die exakte Kopie der DNA. Sie beinhaltet demnach kodierende und nicht kodierende Bereiche. Während der Proteinbiosynthese werden die kodierenden Sequenzen, auch Exons genannt, in eine Aminosäuresequenz überschrieben, wohingegen die nicht kodierenden Bereiche (Introns) herausgeschnitten werden. Wie genau solch ein Schneidprozess (Spleissen) abläuft, ist Gegenstand der vorliegenden Arbeit.

Sc.ai5γ befindet sich im mitochondrialen Genom der Bäckerhefe und gehört zu einer Gruppe von Introns, die sich selbst autokatalytisch spleissen. Es gibt Gemeinsamkeiten in Struktur und Mechanismus zwischen den selbst-spleissenden Introns und den spleissosomalen Introns im Zellkern. Aus diesem Grund glaubt man, dass beide Intronarten gemeinsame evolutionäre Vorfahren haben, was die Untersuchungen am *Sc*.ai5γ besonders interessant macht. Das Gruppe II Intron *Sc*.ai5γ stellt ein ausgezeichnetes Modellsystem dar, indem Faltung und Aktivität untrennbar miteinander verbunden sind. Der Spleissprozess umfasst die RNA Faltung, das anschliessende Spleissen des Introns und das exakte Zusammensetzen der beiden flankierenden Exons. Von allen Introns gehört *Sc*.ai5γ zu dem am besten verstandenen. Trotzdem ist der komplizierte und mehrstufige Spleissprozess dieser grossen dynamischen RNA nur schwer fassbar.

Intronspleissen kann über zwei Wege ablaufen, dem Verzweigungsweg (branching) und dem hydrolytischen Weg. Der Verzweigungsweg hat seinen Namen von dem katalytisch wichtigen Nukleotid, welches sich in einer Zweigstelle in Domäne D6 des Introns befindet. Dieses Nukleotid iniziiert die erste Reaktion im Spleissprozess, das Abspalten des 5'Exons. Im hydrolytischen Weg übernimmt diese Rolle ein Wassermolekül. In beiden Fällen kommt es zum Ligieren der beiden Exons und dem Herausschneiden des Introns, einmal in Form eines Lassos (Verzweigungsweg) oder als lineares Intron (hydrolytischer Weg). Es ist bekannt, dass beide Wege *in vivo* sowie *in vitro* vorkommen. Allerdings sind am Spleissen *in vivo* noch weitere Hilfsproteine beteiligt, während man das Spleissen *in vitro* durch hohe Salzkonzentrationen und erhöhter Temperatur induziert.

Eine faszinierende Eigenschaft von Gruppe II Introns ist, dass man sie nach dem Baukastenprinzip auseinandernehmen und einzelne Domänen wieder zusammenfügen kann. In diesem Zusammenhang entstand das konstruierte Intronderivat D135, das aus drei, statt den sechs original

Domänen besteht. Mit Hilfe dieses Konstrukts untersuchte man den ersten Schritt des hydrolytischen Spleissens, dem die RNA Faltung voran geht. Um die Faltung von D135 mittels Einzelmolekülfluoreszenzspektroskopie untersuchen zu können, musste die RNA fluoreszent markiert werden. Dazu wurden zwei künstliche Schleifen in das Intron eingefügt, L1, L4 (D135-L14, L steht für englisch loop). Die Stellen, an denen sich die Modifikationen befinden, also Domäne D1 und D4, sind höchst -wahrscheinlich nicht direkt an dem Faltungsprozess oder der Katalyse beteiligt (FIGUR 5.1). DNA Oligonukleotide, die die entsprechenden Farbstoffe tragen, hybridisieren an diese Schleifen und markieren so die RNA. Die RNA - Faltung wurde mittels Einzelmolekülmessungen über den Förster - Resonanzenergietransfer (FRET) ermittelt. Einzelmolekül-FRET ist eine geeignete Methode, um den Faltungsprozess von D135-L14 zu untersuchen, da sie Aufschlüsse über ein herterogenes System gibt. Selten vorkommende Zustände oder kleine Subpopulationen können damit aufgedeckt werden. Tatsächlich haben Untersuchungen mittels Einzelmolekül-FRET von D135-L14 fünf Faltungszustände offengelegt, statt der ursprünglichen drei . Demnach faltet sich die RNA, beginnend mit einem ungefalteten Zustand U, über drei partiell gefaltete Zwischenzustände (I_1, I_2, und *F*) zu einer kompakten nativen Struktur. Die Anwesenheit hoher Magnesiumkonzentration (> 40 mM Mg^{2+}), das Exon als Substrat oder das Hilfsprotein begünstigen dabei die Bildung des nativen Zustands. Die Faltung der ersten Domäne D1 ist der limitierende Schritt bei der RNA Faltung und bildet das Gerüst, an das die anderen Domänen andocken. D1 bildet wichtige tertiäre Kontakte mit den beiden flankierenden Exons über die sogenannten Intron-Exon-Bindestellen (FIGUR 5.1). Bei bisherigen Faltungsstudien mittels Einzelmolekül-FRET wurden jedoch die flankierenden Exons ausser Acht gelassen.



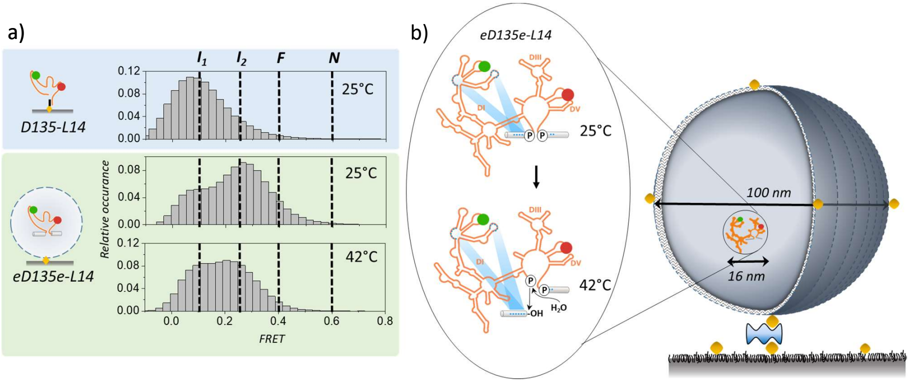
Figur 5.1 Die Sekundärstruktur von eD135e-L14. Die tertiären Kontakte zwischen dem Intron (EBS 1-3) und seinen Exons (IBS 1-3) sind in verschiedenen Blautönen hervorgehoben eD135e-L14 wurde an den Schleifen L1 und L4 floureszent markiert, sowohl mit DNAs, als auch mit PNAs und an den Exons ausschliesslich mit PNAs.

In der vorliegenden Arbeit wird der Einfluss der Exons auf den ersten Schritt des Spleissens, der eine vorangegangene Faltung voraussetzt, untersucht. Wir entwickelten ein neues Konstrukt eD135e-L14, welches die flankierenden Exons trägt und wandten eine neue Strategie zum Markieren von RNA an, mittels PNAs. PNAs sind synthetisch hergestellte Nachbildungen von Nukleinsäuren, bestehend aus den gleichen Basen, aber verbunden mit einer Peptidhauptkette. PNAs haben einige Vorteile gegenüber DNAs: PNAs binden RNA mit einer höheren Affinität und Spezifität. Dank der Verwendung von PNAs erreicht man eine höhere Markierungseffizienz und kann die Grösse der eingefügten Modifikationen zum Markieren von RNA verkürzen. In dieser Studie benutzen wir beides, DNAs und PNAs, um eD135e-L14 in den Schleifen L1 und L4 zu markieren und ausschliesslich PNAs, um die Exons zu markieren.

5.1 Untersuchungen der RNA - Faltung mittels Einzelmolekül-FRET von in Vesikeln eingekapselter RNA

Das KAPITEL II umfasst die Faltungsstudien von eD135e-L14, fluoreszent markiert in den Schleifen L1 und L4, mittels Einzelmolekül-FRET. Vorangegangene Aktivitätsmessung von eD135e-L14 ergaben eine zeitabhängige Abspaltung des 5'Exons und eine zusätzliche Abspaltung des 3'Exons, welche um eine Grössenordnung langsamer abläuft. Weil beide Exons am Faltungsprozess beteiligt sind und abgespalten werden, ist eine direkte Immobilisierung des Konstruktes über die Exons auf der Oberfläche nicht möglich. Eine Alternative bietet die Einkapselung der RNA in Vesikeln, welche ihrerseits an die Oberfläche binden. Auf diesem Weg kann die RNA sich frei bewegen und ihre Faltung ist sterisch nicht behindert. Wir zeigen hier zum ersten Mal die erfolgreiche Einkapselung einer grossen katalytisch aktiven RNA. Mit dieser Anwendung war es uns möglich, den Einfluss der Exons auf die RNA Faltung zu untersuchen.

Vorangegangende Einzelmolekül-FRET Studien von D135-L14 zeigten eine Stabilisierung des nativen Zustandes *N* in Anwesenheit des Substrates. Die Untersuchungen von eD135e-L14, das Intron, welches die kovalent gebundenen Exons trägt, ergaben auch eine Stabilisierung der Faltung, jedoch nicht des nativen Zustands *N* , sondern der Zwischenzustände I_1 und I_2 (FIGUR 5.2).



Figur 5.2 Einzelmolekül-FRET Histogramme von D135-L14 und eD135e-L14 und das Schema der Einkapselung. ***a)*** *Einzelmolekül-FRET Histogramme von D135-L14, welches direkt immobilisiert wurde bei25 °C, im Vergleich zum eingekapselten Konstrukt eD135e-L14 bei 25 °C und 42 °C. Die Faltungzustände I1, I2, F und N sind in jedem Histogramm angegeben.* ***b)*** *Das Schema der Einkapselung von eD135e-L14 zeigt die RNA im Vesikel. Diese Methode ermöglicht die Aktivierung der RNA mittels Temperaturerhöhung.*

Der nächste Schritt beschreibt die Aktivierung der RNA mittels Temperaturerhöhung. Alle zuvor durchgeführten Einzelmolekül-FRET-Studien wurden bei 25 °C durchgeführt. Dank den Aktivitätsmessungen wissen wir aber, dass bei dieser Temperatur keine Abspaltung des 5'Exons geschieht. Wir induzierten das Spleissen mittels Temperatur auf dem Mikroskop in den eingekapselten RNA-Molekülen (FIGUR 5.2). Unsere Ergebnisse zeigen, dass das abgespaltene 5'Exon am Intron gebunden bleibt. Nach anschliessendem Abkühlen der Probe auf 25 °C nimmt die RNA wieder die gleichen Faltungszustände ein wie vor der Aktivierung. Wir sehen auch Unterschiede bei der Verwendung der beiden Proben DNA und PNA. Diese Unterschiede sind auf die flexibleren Schleifen mit PNAs zurückzuführen. Die kürzeren PNAs binden nur an die Hälfte der Schleifen.

Wir konnten zeigen, dass die Einkapselung von RNA eine vielversprechende Methode ist, um katalytisch aktive und frei diffundierende RNA mittels Einzelmolekül-FRET zu untersuchen. Mit Hilfe dieser Technik gelang es uns, die Rolle der Exons während der RNA Faltung zu untersuchen.

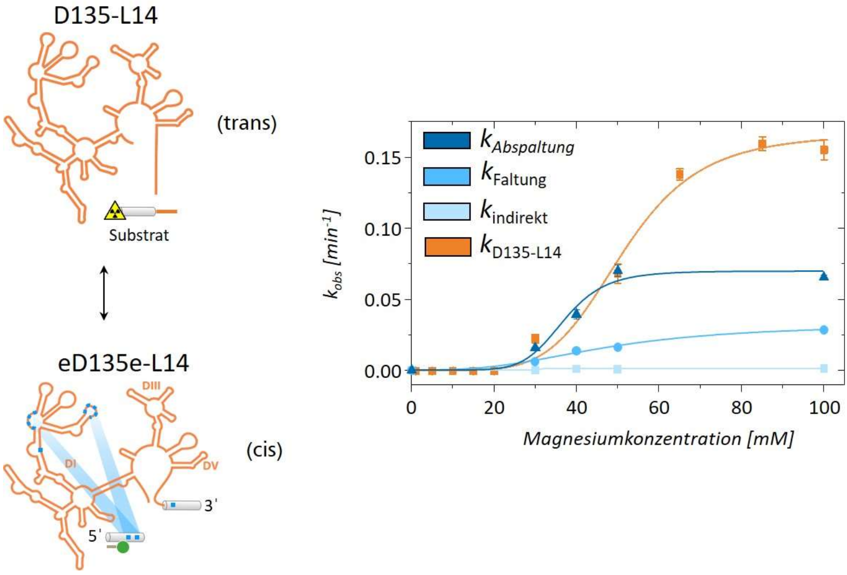
5.2 Faltungs- und Abspaltungskinetic von eD135e-L14

Der Energietransfer zwischen zwei Farbstoffen ist abhängig von ihrer Distanz zueinander. FRET wird auch als molakulares Linear verwendet. Die Distanz zweier Farbstoffe gibt Aufschlüsse über Struktur des Biomoleküls, an dem sie angeheftet sind. Wir benutzen eine modulierte dreidimensionale Struktur eines Introns (eD1356), welches aus einem kurzen 5'Exon und den Domänen D1, D3, D5 und D6 besteht, um theoretische FRET-Werte zu bestimmen, die mit den Farbstoffpositionen in eD135e-L14 übereinstimmen. Dabei fanden wir eine gute Übereinstimmung zwischen dem berechneten FRET-Wert von eD1356 mit dem experimentell bestimmten Zwischenzustand I_1 von eD135e-L14. Dieses Ergebnis deutet daraufhin, dass der erste Zwischenzustand genau der Konformation der Modellstruktur entspricht. Alle höheren experimentell bestimmten FRET-Zustände, wie I_2, *F* und *N,* weisen entweder auf eine noch kompaktere Struktur hin, oder entstehen durch die Flexibilität der künstlich eingefügten Schleifen.

Die zwei zusätzlich eingefügten Exons bieten weitere Möglichkeiten, die RNA eD135e-L14 fluoreszent zu markieren. In KAPITEL III wird der Einfluss der Exons mit Hilfe der zusätzlichen Markierungen noch detaillierter untersucht. Wir vergleichen dazu die experimentell bestimmten FRET-Zustände mit den berechneten FRET- Werten der Modellstruktur. Unsere Ergebnisse zeigen einen hochdynamischen reversiblen Faltungsweg, in dem die Modellstruktur einen Zwischenzustand darstellt. Wir beobachteten kompaktere RNA-Strukturen , als die Modellstruktur vorgibt. Darüberhinaus können wir nun den FRET-Zustand Null aufklären und ihn eindeutig dem Faltungsweg zuordnen. Die Änderungen der Population des FRET- Zustands Null zeigen, dass es sich hierbei nicht um ein Artefakt handelt, sondern eher eine ungefaltete RNA-Struktur darstellt, die langsam in einen gefalteten Zustand übergehen kann.

Im zweiten Teil des Kapitels geht es um den ersten Schritt des Spleissen, nämlich das Abspalten des 5'Exons vom Intron. Die Kinetiken dieser Reaktion haben wir mit sogenannten fluoreszenten nativen Gelen durchgeführt. Dabei wurde die RNA eD135e-L14 an den flankierenden Exons mittels PNAs markiert. Ein Vergleich des gleichen Experimentes mit radioaktiv markierter RNA ergab, dass die PNAs die Aktivität der RNA nicht einschränken. Aus diesem Grund wurden alle weiteren

Experimente mit den PNAs durchgeführt. Die Aktivitätsbestimmung von eD135e-L14 half dabei, die drei sehr eng miteinander verbunden Prozesse zu entflechten; RNA- Faltung, Abspaltung des Exons und einen indirekten Faltungszustand. Letzterer beschreibt eine missgefaltete Struktur, die über die Zeit in den aktiven Faltungszustand übergeht. Alle drei Prozesse lassen sich sehr gut mit dem Modell der Serienreaktion mit indirektem Faltungszustand beschreiben. Der Vergleich mit den Literaturwerten, bei denen ein Substrat in *trans* zu dem Konstrukt D135-L14 dazugegeben wird und unseren Studien, bei denen sich das Exon in *cis* am Intron befindet, zeigen deutliche Unterschiede (FIGUR 5.3). Beide Reaktionen unterscheiden sich in der Umsatzrate und in den Geschwindigkeitsraten, haben aber gleiche Dissoziationskonstanten für Magnesium (K_{Mg}). Ein weiterer Vergleich zwischen *Sc.*ai5γ mit kurzen Exons und eD135e-L14 zeigt gleiche Geschwindigkeitsraten, wohingegen die aller anderen Derivate viel schneller sind. Dieses Ergebnis deutet darauf hin, dass die Faltung der limitierende Schritt während des Spleissens ist.



Figur 5.3 Ein Vergleich zwischen D135-L14 und seinem Substrat, was in trans dazugegeben wird, und dem Exon-tragenden Konstrukt eD135e-L14. Die Geschwindigkeitskonstanten KAbspaltung, kFaltung, und kindirekt ergeben sich aus dem Fit-Model der Serienreaktion mit indirektem Faltungszustand. Hingegen wurde die Substratspaltung von D135-L14 nur mit einem exponentiellen Modell bestimmt und ergab eine Geschwindigkeitskonstante.

Appendix for chapter II

Table 2.13 Summary of all conditions and the percentage of dynamic, static and zero FRET and the total number of molecules. If not in particular noted, then the standard conditions (DNA labels, exposure time of 28 ms, vesicle encapsulation) were applied. All smFRET experiments were carried out in the presence of 500 mM KCl and 100 mM $MgCl_2$*, pH 6.9.*

Condition	**Percentage of molecules**			**Total number of molecules**
	Dynamic FRET	Static FRET	Zero FRET	
D135-L14 surface immobilized	0.66	0.25	0.09	143
D135-L14 + 3'-linker encapsulated 25 °C	0.43	0.19	0.38	194
D135-L14 + 3'-linker encapsulated 42 °C	0.34	0.33	0.33	132
D135-L14 + 3'-linker encapsulated cool down 25 °C	0.54	0.18	0.29	28
D135-L14 encapsulated 25 °C	0.66	0.16	0.18	89
D135-L14 encapsulated 42 °C	0.47	0.37	0.16	49
D135-L14 encapsulated cool down 25 °C	0.48	0.36	0.17	42
eD135e-L14 25 °C	0.51	0.32	0.17	198
eD135e-L14 42 °C	0.59	0.14	0.27	49
eD135e-L14 cool down 25 °C	0.45	0.35	0.20	174
eD135e-L14 25 °C 100 ms	0.60	0.30	0.10	87
eD135e-L14 25 °C 28 ms	0.55	0.36	0.09	120
eD135e-L14 25 °C 14 ms	0.50	0.39	0.11	118
eD135e-L14 25 °C 2 ms	0.56	0.33	0.11	66
eD135e-L14 25 °C PNAs	0.57	0.35	0.08	84
eD135e-L14 42 °C PNAs	0.34	0.58	0.08	466

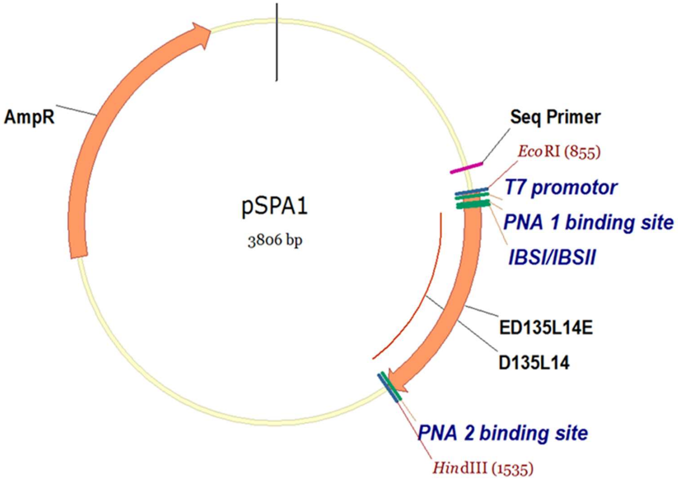


Figure 2.60 Plasmid map of pSPA1 (3806 bp) which was used as a template to transcribe the RNA construct eD135e-L14. The gene for the ampicillin resistance (Amp) is shown in orange. The insert eD135e-L14 is highlighted in orange. The restriction sites EcoRI and HindIII were used for cloning.

DNA sequence of the insert eD135e-L1 from 5' to 3'-direction. The sequence in black refers to the intron, while the flanking exons are depicted in dark blue. The sequence highlighted in yellow indicates the PNA binding sites and the bolt letters refers to the intron-binding sites 1, 2 and 3.

GGGGAGACTTACTAC**GTGGTG**GGA**CATTTTC***GAGCGGTCTGAAAGTTATCATAAATAATATTTACCATATAATAATGGATAAATTATATTTTTATCAATATAAGTCTAATTACAAGTGTATTAAAATGGTAACATAAATATGCTAAGCTGTAATGACAAAAGTATCCATATTCTTGACAGTTATTTTATATTATAAAAAAAAGATGAAGGAACTTTGACTGATCTAATATGCTCAACGAAAGTGAATCAAATGTTATAAAATTACTTACACCACTAATTGAAAACCTGTCTGATATTCAATTATTATCCCAAATTAATAACGCTCTTGGTAGGGATAAAATGGTTGATGTTATGTATTGGAAATGAGCATACGATAAATCATATAACCATTAGTAATATAATTTGAGAGCTAAGTTAGATATTTACGTATTTATGATAAAACAGAATAAACCCTCTTCGGAGGGTAAAAGATTGTATAAAAAGCTAATGCCATATTGTAATGATATGGATAAGAATTATTATTCTAAAGATGAAAATCTGCTAACTTATACTATAGGTGATACCCAAATATATTTCGGACTACGTAGGGTATTGCGTGAGCCGTATGCGATGAAAGTCGCACGTACGGTTCCTGGAT***A**CTATGTATTATCA

Appendix for chapter III

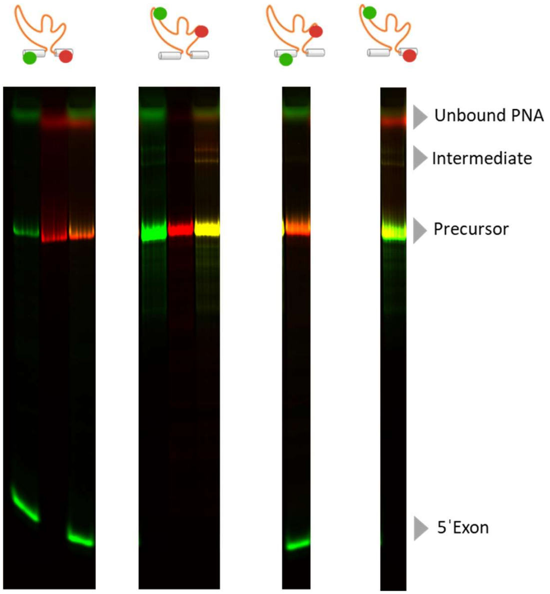


Figure 3.61 5 % native PAGE of eD135e-L14 labeled fluorescently at different positions. From left to right; eD135e-L14 labeled at the flanking exons, eD135e labeled at the internal loops L1 and L4, eD135e-L14 labeled at the 5'exon and L4, and eD135e-L14 labeled at the 3'exon and L1. The splicing was performed under standard conditions 500 Mm KCl, 100 Mm $MgCl_2$ at Ph 6.9 AND 42 °C for 60 min. Shown is an overlay of Cy3 and Cy5 channel. A co-localization of Cy3 (green) and Cy5 (red) gives a yellow band.

Table 3.14 Summary of all different labeling positions for eD135e-L14 and the percentage of dynamic, static and zero FRET and the total number of molecules. All smFRET experiments were carried out in the presence of 500 mM KCl and 100 mM $MgCl_2$, pH 6.9.

Condition		Percentage of molecules			Total number of molecules
		Dynamic FRET	Static FRET	Zero FRET	
DNA L1/DNA L4	25 °C	0.51	0.32	0.17	198
DNA L1/DNA L4	42 °C	0.59	0.14	0.27	49
DNA L1/DNA L4 cool down	25 °C	0.45	0.35	0.20	174
PNA L1/PNA L4	25 °C	0.57	0.35	0.08	84
PNA L1/PNA L4	42 °C	0.34	0.58	0.08	466
PNA 5'E/PNA 3'E	25 °C	0.36	0.11	0.53	87
PNA 5'E/PNA 3'E	42 °C	0.47	0.04	0.48	118
PNA 5'E/PNA L4	25 °C	0.50	0.20	0.30	140
PNA 5'E/PNA L4	42 °C	0.65	0.20	0.16	215
PNA 5'E/DNA L1	25 °C	0.50	0.42	0.08	153
PNA 5'E/DNA L1	42 °C	0.63	0.35	0.02	102
PNA 3'E/DNA L1	25 °C	0.18	0.11	0.70	368
PNA 3'E/DNA L1	42 °C	0.36	0.13	0.50	214
PNA 3'E/DNA L1 cool down	25 °C	0.35	0.26	0.38	34

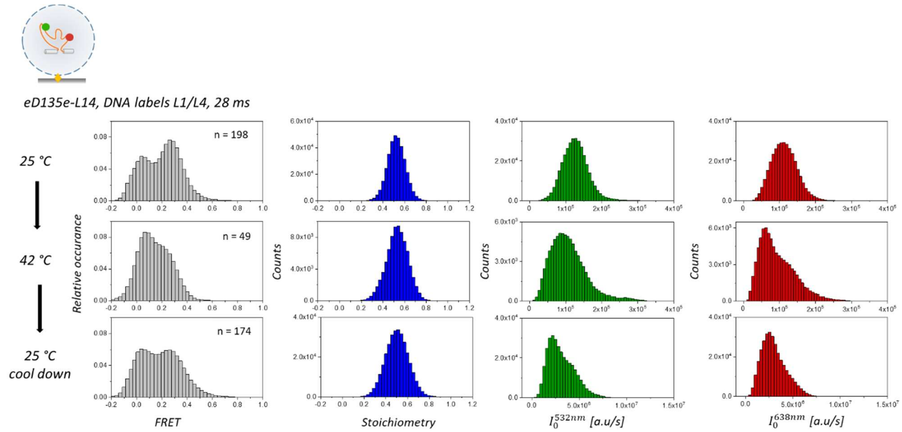


Figure 3.62 Summary of all detected smFRET signals by investigating the folding of the encapsulated eD135e-L14 labeled with DNA-Cy3 in L1 and with DNA-Cy5 in L1. From left to right, the following histograms are shown: FRET, stoichiometry, the total fluorescent intensity upon green and red at 25 °C (top), 42 °C (middle), and 25 °C cool down (bottom). The number of molecules (n) represents dynamic, static and zero FRET molecules.

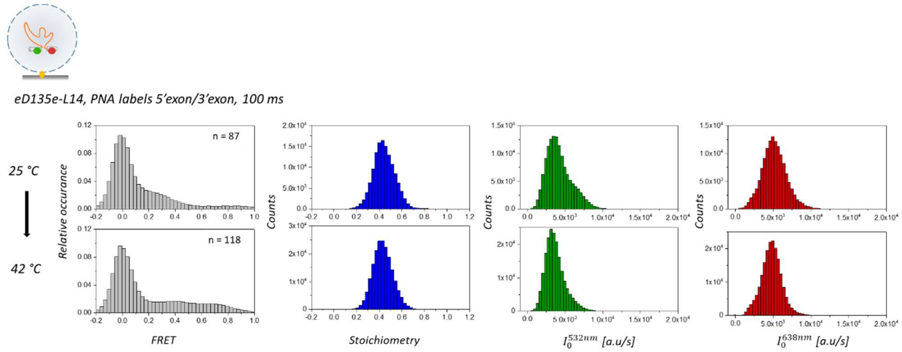


Figure 3.63 Summary of all detected smFRET signals by investigating the folding of the encapsulated eD135e-L14 labeled with PNA-Cy3 at the 5'exon and with PNA-Cy5 at the 3'exon. From left to right, the following histograms are shown: FRET, stoichiometry, the total fluorescent intensity upon green and red at 25 °C (top) and 42 °C (bottom). The number of molecules (n) represents dynamic, static and zero FRET molecules.

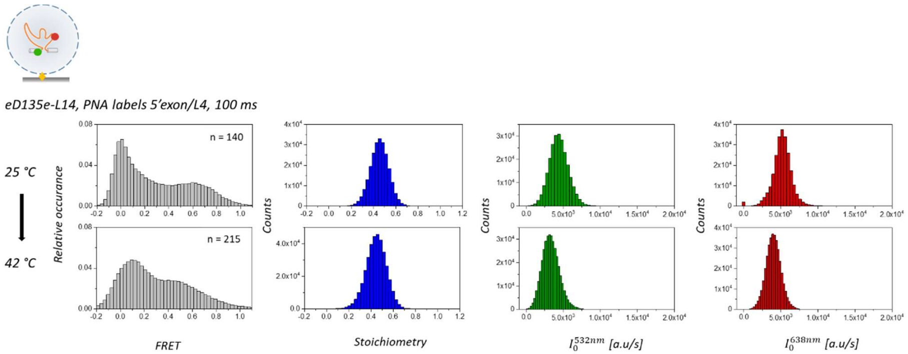


Figure 3.64 Summary of all detected smFRET signals by investigating the folding of the encapsulated eD135e-L14 labeled with PNA-Cy3 at the 5'exon and with DNA-Cy5 in L4. From left to right, the following histograms are shown: FRET, stoichiometry, the total fluorescent intensity upon green and red at 25 °C (top) and 42 °C (bottom). The number of molecules (n) represents dynamic, static and zero FRET molecules.

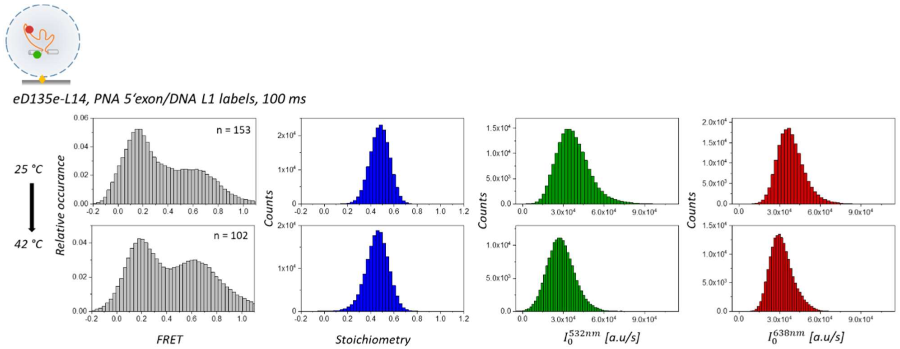


Figure 3.65 Summary of all detected smFRET signals by investigating the folding of the encapsulated eD135e-L14 labeled with PNA-Cy3 at the 5'exon and with DNA-Cy5 in L1. From left to right, the following histograms are shown: FRET, stoichiometry, the total fluorescent intensity upon green and red at 25 °C (top) and 42 °C (bottom). The number of molecules (n) represents dynamic, static and zero FRET molecules.

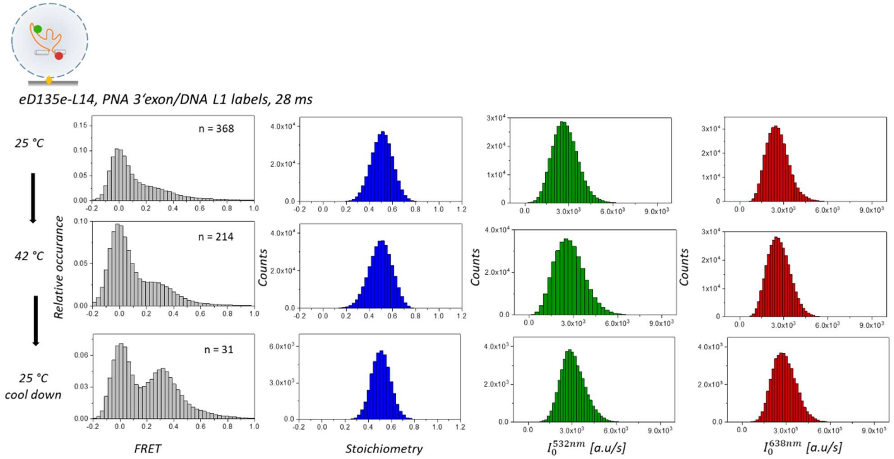


Figure 3.66 Summary of all detected smFRET signals by investigating the folding of the encapsulated eD135e-L14 labeled with PNA-Cy5 at the 3'exon and with DNA-Cy3 in L1. From left to right, the following histograms are shown: FRET, stoichiometry, the total fluorescent intensity upon green and red at 25 °C (top), 42 °C (middle) and 25 °C cool down (bottom). The number of molecules (n) represents dynamic, static and zero FRET molecules.

Table 3.15 Determined rate constants for the cis-cleavage of eD135e-L14 applying the single exponential growth model (Paragraph 3.4.7 (i)) for 40 and 100 mM $MgCl_2$.

$MgCl_2$ [mM]	k [min^{-1}]
40	$1.1^{-2} \pm 9.3^{-4}$
100	$1.8^{-2} \pm 2.0^{-3}$

Table 3.16 Determined rate constants for the cis-cleavage of eD135e-L14 applying the series reaction model (global fit) model (Paragraph 3.4.7 (ii)) for 40 and 100 mM $MgCl_2$.

$MgCl_2$ [mM]	k_1 [min^{-1}]	k_2 [min^{-1}]
40	$1.4^{-2} \pm 4.6^{-4}$	$8.9^{-2} \pm 5.4^{-3}$
100	$2.4^{-2} \pm 5.1^{-4}$	$1.3^{-1} \pm 8.9^{-3}$

Table 3.17 Comparison of rate constants between the cis-cleavage of eD135e-L14 and the trans-cleavage of D135-L14. The rate constants for eD135e-L14 were determined using the series reaction model (global fit) including an off-pathway (Paragraph 3.4.7 (iii)). The cleavage reaction is described by k_1, the RNA folding by k_2, and the off-pathway by k_3. The activity of D135-L14 was determined in the presence of the DNA labels and was performed by Dr. Erica Fiorini (Fiorini, 2015). The assay condition was 500 mM KCl, pH 6.9, 42 °C and the specified Mg^{2+} concentrations.

$MgCl_2$ [mM]	eD135e-L14			D135-L14
	k_1 [min^{-1}]	k_2 [min^{-1}]	k_3 [min^{-1}]	k_{obs} [min^{-1}]
30	$1.6^{-2} \pm 5.4^{-4}$	$5.9^{-3} \pm 1.0^{-4}$	$4.0^{-5} \pm 9.2^{-5}$	$2.0^{-2} \pm 3.1^{-3}$
40	$3.9^{-2} \pm 3.1^{-3}$	$1.4^{-2} \pm 4.6^{-4}$	$1.2^{-3} \pm 4.6^{-4}$	-
50	$7.0^{-2} \pm 4.2^{-3}$	$1.6^{-2} \pm 3.4^{-4}$	$7.5^{-4} \pm 3.8^{-4}$	$6.8^{-2} \pm 7.0^{-3}$
100	$6.6^{-2} \pm 1.6^{-3}$	$2.8^{-2} \pm 3.4^{-4}$	$1.2^{-3} \pm 3.8^{-4}$	$1.5^{-1} \pm 7.0^{-3}$

6. References

Armitage, B.A. (2011), “Imaging of RNA in live cells”, *Current Opinion in Chemical Biology*, Vol. 15 No. 6, pp. 806–812.

Belhocine, K., Mak, A.B. and Cousineau, B. (2007), “Trans-splicing of the Ll.LtrB group II intron in Lactococcus lactis”, *Nucleic Acids Research*, Vol. 35 No. 7, pp. 2257–2268.

Börner, R., Kowerko, D., Miserachs, H.G., Schaffer, M.F. and Sigel, R.K.O. (2016), “Metal ion induced heterogeneity in RNA folding studied by smFRET”, *Coordination Chemistry Reviews*.

Boukobza, E., Sonnenfeld, A. and Haran, G. (2001), “Immobilization in Surface-Tethered Lipid Vesicles as a New Tool for Single Biomolecule Spectroscopy”, *Journal of Physical Chemistry B*, Vol. 105 No. 48, pp. 12165–12170.

Cech, T.R. (1971), “The genetic organization of chromosomes”, *Annual Review of Genetics*, Vol. 5, pp. 237–256.

Cech, T.R. (1986), “The generality of self-splicing RNA. Relationship to nuclear mRNA splicing”, *Cell*, Vol. 44 No. 2, pp. 207–210.

Chan, R.T., Robart, A.R., Rajashankar, K.R., Pyle, A.M. and Toor, N. (2012), “Crystal structure of a group II intron in the pre-catalytic state”, *Nature Structural Molecular Biology*, Vol. 19 No. 5, pp. 555–557.

Chandradoss, S.D., Haagsma, A.C., Lee, Y.K., Hwang, J.-H., Nam, J.-M. and Joo, C. (2014), “Surface passivation for single-molecule protein studies”, *Journal of Visualized Experiments*, No. 86.

Chanfreau, G. and Jacquier, A. (1994), “Catalytic site components common to both splicing steps of a group II intron”, *Science*, Vol. 266 No. 5189, pp. 1383–1387.

Chinnery, P.F., Taylor, R.W., Diekert, K., Lill, R., Turnbull, D.M. and Lightowlers, R.N. (1999), “Peptide Nucleic Acid Delivery To Human Mitochondria”, *Gene Therapy*, Vol. 6 No. 12, pp. 1919–1928.

Cisse, I., Okumus, B., Joo, C. and Ha, T. (2007), “Fueling protein DNA interactions inside porous nanocontainers”, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 104 No. 31, pp. 12646–12650.

Cisse, I.I., Kim, H. and Ha, T. (2012), “A rule of seven in Watson-Crick base-pairing of mismatched sequences”, *Nature Structural Molecular Biology*, Vol. 19 No. 6, pp. 623–627.

Cordes, T., Vogelsang, J. and Tinnefeld, P. (2009), “On the mechanism of Trolox as antiblinking and antibleaching reagent”, *Journal of the American Chemical Society*, Vol. 131 No. 14, pp. 5018–5019.

Costa, M., Michel, F. and Westhof, E. (2000), “A three-dimensional perspective on exon binding by a group II self-splicing intron”, *The EMBO Journal*, Vol. 19 No. 18, pp. 5007–5018.

Costa, M., Walbott, H., Monachello, D., Westhof, E. and Michel, F. (2016), “Crystal structures of a group II intron lariat primed for reverse splicing”, *Science*, Vol. 354 No. 6316.

Dai, L., Toor, N., Olson, R., Keeping, A. and Zimmerly, S. (2003), “Database for mobile group II introns”, *Nucleic Acids Research*, Vol. 31 No. 1, pp. 424–426.

Dai, L. and Zimmerly, S. (2002), “Compilation and analysis of group II intron insertions in bacterial genomes. Evidence for retroelement behavior”, *Nucleic Acids Research*, Vol. 30 No. 5, pp. 1091–1102.

Daniels, D.L., Michels, W.J., JR. and Pyle, A.M. (1996), “Two Competing Pathways for Self-splicing by Group II Introns: A Quantitative Analysis ofin VitroReaction Rates and Products”, *Journal of Molecular Biology*, Vol. 256 No. 1, pp. 31–49.

Dimura, M., Peulen, T.O., Hanke, C.A., Prakash, A., Gohlke, H. and Claus, S. (2016), “Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems”, *Current Opinion in Structural Biology*, Vol. 40, pp. 163–185.

Dlakić, M. and Mushegian, A. (2011), “Prp8, the pivotal protein of the spliceosomal catalytic center, evolved from a retroelement-encoded reverse transcriptase”, *RNA*, Vol. 17 No. 5, pp. 799–808.

Donghi, D., Pechlaner, M., Finazzo, C., Knobloch, B. and Sigel, R.K.O. (2013), “The structural stabilization of the κ three-way junction by Mg(II) represents the first step in the folding of a group II intron”, *Nucleic Acids Research*, Vol. 41 No. 4, pp. 2489–2504.

Egholm, M., Buchardt, O., Christensen, L., Behrens, C., Freier, S.M., Driver, D.A., Berg, R.H., Kim, S.K., Norden, B. and Nielsen, P.E. (1993), “PNA Hybridizes To Complementary Oligonucleotides Obeying The Watson-Crick Hydrogen-Bonding Rules”, *Nature*, Vol. 365 No. 6446, pp. 566–568.

Egloff, D., Oleinich, I.A., Zhao, M., König, S.L.B., Sigel, R.K.O. and Freisinger, E. (2016), “Sequence-Specific Post-Synthetic Oligonucleotide Labeling for Single-Molecule Fluorescence Applications”, *ACS Chemical Biology*, Vol. 11 No. 9, pp. 2558–2567.

Erat, M. and Sigel, R.K.O. (2008), “Divalent metal ions tune the self-splicing reaction of the yeast mitochondrial group II intron Sc.ai5γ”, *Journal of Biological Inorganic Chemistry*, Vol. 13 No. 6, pp. 1025–1036.

Erat, M.C., Zerbe, O., Fox, T. and Sigel, R.K.O. (2007), “Solution structure of domain 6 from a self-splicing group II intron ribozyme: a Mg(2+) binding site is located close to the stacked branch adenosine”, *Chembiochem: A European Journal of Chemical Biology*, Vol. 8 No. 3, pp. 306–314.

Farooq, S. and Hohlbein, J. (2015), “Camera-based single-molecule FRET detection with improved time resolution”, *Physical Chemistry Chemical Physics*, Vol. 17 No. 41, pp. 27862–27872.

Fazliji, B. (2017), “Following the splicing process of an encapsulated group II intron by single-molecule FRET”, Master thesis, University of Zurich, Switzerland, 2017.

Fedorova, O. (2012), “Kinetic characterization of group II intron folding and splicing”, *Methods in Molecular Biology*, Vol. 848, pp. 91–111.

Fedorova, O., Mitros, T. and Pyle, A.M. (2003), “Domains 2 and 3 interact to form critical elements of the group II intron active site”, *Journal of Molecular Biology*, Vol. 330 No. 2, pp. 197–209.

Fedorova, O. and Pyle, A.M. (2012), “The brace for a growing scaffold: Mss116 protein promotes RNA folding by stabilizing an early assembly intermediate”, *Journal of Molecular Biology*, Vol. 422 No. 3, pp. 347–365.

Fedorova, O., Waldsich, C. and Pyle, A.M. (2007), “Group II intron folding under near-physiological conditions. Collapsing to the near-native state”, *Journal of Molecular Biology*, Vol. 366 No. 4, pp. 1099–1114.

Fedorova, O. and Zingler, N. (2007), “Group II introns: structure, folding and splicing mechanism”, *Biological Chemistry*, Vol. 388 No. 7.

Fica, S.M., Tuttle, N., Novak, T., Li, N.-S., Lu, J., Koodathingal, P., Dai, Q., Staley, J.P. and Piccirilli, J.A. (2013), “RNA catalyses nuclear pre-mRNA splicing”, *Nature*, Vol. 503 No. 7475, pp. 229–234.

Fiorini, E. (2015), “From Bulk to Single Molecule: Crowding Agents and Mutation Effects on Activity and Folding of a Group II intron Ribozyme”, Ph.D. thesis, University of Zurich, 2015.

Fiorini, E., Börner, R. and Sigel, R.K.O. (2015), “Mimicking the *in vivo* Environment – The Effect of Crowding on RNA and Biomacromolecular Folding and Activity”, *CHIMIA International Journal for Chemistry*, Vol. 69 No. 4, pp. 207–212.

Fish, K.N. (2009), “Total internal reflection fluorescence (TIRF) microscopy”, *Current Protocols in Cytometry*, Chapter 12, Unit12.18.

Galej, W.P., Oubridge, C., Newman, A.J. and Nagai, K. (2013), “Crystal structure of Prp8 reveals active site cavity of the spliceosome”, *Nature*, Vol. 493 No. 7434, pp. 638–643.

Gallo, S., Furler, M. and Sigel, R.K.O. (2005), “In vitro transcription and purification of RNAs of different size”, *Chimia*, Vol. 59 No. 11, pp. 812–816.

Gilbert, W. (1978), “Why Genes In Pieces”, *Nature*, Vol. 271 No. 5645, p. 501.

Gregory, T.R. (2001), “Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma”, *Biological reviews of the Cambridge Philosophical Society*, Vol. 76 No. 1, pp. 65–101.

Griffin, E.A., JR, Qin, P.Z., Michels, W.J., JR. and Pyle, A.M. (1995), “Group II intron ribozymes that cleave DNA and RNA linkages with similar efficiency, and lack contacts with substrate 2'-hydroxyl groups”, *Chemistry and Biology*, Vol. 2 No. 11, pp. 761–770.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1983), “The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme”, *Cell*, Vol. 35 No. 3, pp. 849–857.

Ha, T. (2001), “Single-molecule fluorescence methods for the study of nucleic acids”, *Current Opinion in Structural Biology*, Vol. 11 No. 3, pp. 287–292.

Ha, T. (2007), “The New Era of Biology In Singulo”, *Cold Spring Harbor Laboratory Press*.

Ha, T., Enderle, T., Ogletree, D.F., Chemla, D.S., Selvin, P.R. and Weiss, S. (1996), “Probing the interaction between two single molecules. Fluorescence resonance energy transfer between a single donor and a single acceptor”, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 93 No. 13, pp. 6264–6268.

Hadzic, M.C.A.S. (2017), “Vesicle Encapsulation and Data Analysis Standardization to Characterize Large Catalytic RNAs Using Single Molecule FRET”, Ph.D. thesis, University of Zurich, Switzerland, 2017.

Hadzic, M.C.A.S., Kowerko, D., Börner, R., Zelger-Paulus, S. and Sigel, R.K.O. (2016), “Detailed analysis of complex single molecule FRET data with the software MASH”, in Farkas, D.L., Nicolau, D.V. and Leif, R.C. (Eds.), *San Francisco, California, United States, Saturday 13 February 2016*, SPIE, p. 971119.

Halls, C., Mohr, S., Del Campo, M., Yang, Q., Jankowsky, E. and Lambowitz, A.M. (2007), “Involvement of DEAD-box proteins in group I and group II intron splicing. Biochemical characterization of Mss116p, ATP hydrolysis-dependent and -independent mechanisms, and general RNA chaperone activity”, *Journal of Molecular Biology*, Vol. 365 No. 3, pp. 835–855.

Harvey, B.J. and Levitus, M. (2009), "Nucleobase-specific enhancement of Cy3 fluorescence", *Journal of Fluorescence*, Vol. 19 No. 3, pp. 443–448.

Jarrell, K.A., Peebles, C.L., Dietrich, R.C., Romiti, S.L. and Perlman, P.S. (1988), "Group II intron self-splicing. Alternative reaction conditions yield novel products", *Journal of Biological Chemistry*, Vol. 263 No. 7, pp. 3432–3439.

Jo, B.-S. and Choi, S.S. (2015), "Introns. The Functional Benefits of Introns in Genomes", *Genomics and Informatics*, Vol. 13 No. 4, pp. 112–118.

Kapanidis, A.N., Laurence, T.A., Lee, N.K., Margeat, E., Kong, X. and Weiss, S. (2005), "Alternating-laser excitation of single molecules", *Accounts of chemical research*, Vol. 38 No. 7, pp. 523–533.

Kapanidis, A.N., Majumdar, D., Heilemann, M., Nir, E. and Weiss, S. (2015), "Alternating Laser Excitation for Solution-Based Single-Molecule FRET", *Cold Spring Harbor protocols*, Vol. 2015 No. 11, pp. 979–987.

Karunatilaka, K.S., Solem, A., Pyle, A.M. and Rueda, D. (2010), "Single-molecule analysis of Mss116-mediated group II intron folding", *Nature*, Vol. 467 No. 7318, pp. 935–939.

Klehs, K., Spahn, C., Endesfelder, U., Lee, S.F., Fürstenberg, A. and Heilemann, M. (2014), "Increasing the brightness of cyanine fluorophores for single-molecule and superresolution imaging", *ChemPhysChem: A European Journal of Chemical Physics and Physical Chemistry*, Vol. 15 No. 4, pp. 637–641.

Koch, J.L. (1992), "Group II introns deleted for multiple substructures retain self-splicing activity", *Molecular and Cellular Biology*, No. 12, pp. 1950–1958.

Kowerko, D., König, S.L.B., Skilandat, M., Kruschel, D., Hadzic, M.C.A.S., Cardo, L. and Sigel, R.K.O. (2015), "Cation-induced kinetic heterogeneity of the intron-exon recognition in single group II introns", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 112 No. 11, pp. 3403–3408.

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982), "Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena", *Cell*, Vol. 31 No. 1, pp. 147–157.

Kummer, S., Knoll, A., Socher, E., Bethge, L., Herrmann, A. and Seitz, O. (2011), "Fluorescence imaging of influenza H1N1 mRNA in living infected cells using single-chromophore FIT-PNA", *Angewandte Chemie International Edition*, Vol. 50 No. 8, pp. 1931–1934.

Kummer, S., Knoll, A., Socher, E., Bethge, L., Herrmann, A. and Seitz, O. (2012), “PNA FIT-probes for the dual color imaging of two viral mRNA targets in influenza H1N1 infected live cells”, *Bioconjugate Chemistry*, Vol. 23 No. 10, pp. 2051–2060.

Lambowitz, A.M. and Zimmerly, S. (2011), “Group II introns: mobile ribozymes that invade DNA”, *Cold Spring Harbor perspectives in biology*, Vol. 3 No. 8, pp. a003616.

Lamichhane, R., Solem, A., Black, W. and Rueda, D. (2010), “Single-molecule FRET of protein-nucleic acid and protein-protein complexes: surface passivation and immobilization”, *Methods*, Vol. 52 No. 2, pp. 192–200.

Lander, E.S. et al. (2001), “Initial sequencing and analysis of the human genome”, *Nature*, Vol. 409 No. 6822, pp. 860–921.

Lee, S., Lee, J. and Hohng, S. (2010), “Single-molecule three-color FRET with both negligible spectral overlap and long observation time”, *PloS One*, Vol. 5 No. 8, e12270.

Li, C.-F., Costa, M., Bassi, G., Lai, Y.-K. and Michel, F. (2011), “Recurrent insertion of 5'-terminal nucleotides and loss of the branchpoint motif in lineages of group II introns inserted in mitochondrial preribosomal RNAs”, *RNA*, Vol. 17 No. 7, pp. 1321–1335.

Lifton, R.P., Goldberg, M.L., Karp, R.W. and Hogness, D.S. (1977), “Organization Of Histone Genes In Drosophila-Melanogaster - Functional And Evolutionary Implications”, *Cold Spring Harb Sym*, Vol. 42, pp. 1047–1051.

Liu, B., Mazouchi, A. and Gradinaru, C.C. (2010), “Trapping single molecules in liposomes. Surface interactions and freeze-thaw effects”, *The journal of physical chemistry. B*, Vol. 114 No. 46, pp. 15191–15198.

Marcia, M. and Pyle, A.M. (2012), “Visualizing group II intron catalysis through the stages of splicing”, *Cell*, Vol. 151 No. 3, pp. 497–507.

Marcia, M. and Pyle, A.M. (2014), “Principles of ion recognition in RNA: insights from the group II intron structures”, *RNA*, Vol. 20 No. 4, pp. 516–527.

McCann, J.J., Choi, U.B., Zheng, L., Weninger, K.R. and Bowen, M.E. (2010), “Optimizing methods to recover absolute FRET efficiency from immobilized single molecules”, *Biophys Journal*, Vol. 99 No. 3, pp. 961–970.

McNeil, B.A., Semper, C. and Zimmerly, S. (2016), “Group II introns. Versatile ribozymes and retroelements”, *Wiley interdisciplinary reviews. RNA*, Vol. 7 No. 3, pp. 341–355.

Michels, W.J., JR. and Pyle, A.M. (1995), “Conversion of a Group II Intron into a New Multiple-Turnover Ribozyme that Selectively Cleaves Oligonucleotides. Elucidation of Reaction Mechanism and Structure/Function Relationships”, *Biochemistry*, Vol. 34 No. 9, pp. 2965–2977.

Molina-Sánchez, M.D., Martinez-Abarca, F. and Toro, N. (2006), “Excision of the Sinorhizobium meliloti group II intron RmInt1 as circles in vivo”, *Journal of Biological Chemistry*, Vol. 281 No. 39, pp. 28737–28744.

Mujumdar, R.B., Ernst, L.A., Mujumdar, S.R., Lewis, C.J. and Waggoner, A.S. (1993), “Cyanine dye labeling reagents. Sulfoindocyanine succinimidyl esters”, *Bioconjugate Chemistry*, Vol. 4 No. 2, pp. 105–111.

Muratovska, A., Lightowlers, R.N., Taylor, R.W., Turnbull, D.M., Smith, R. A. J., Wilce, J.A., Martin, S.W. and Murphy, M.P. (2001), “Targeting peptide nucleic acid (PNA) oligomers to mitochondria within cells by conjugation to lipophilic cations. Implications for mitochondrial DNA replication, expression and disease”, *Nucleic Acids Research*, Vol. 29 No. 9, pp. 1852–1863.

Murray, H.L., Mikheeva, S., Coljee, V.W., Turczyk, B.M., Donahue, W.F., Bar-Shalom, A. and Jarrell, K.A. (2001), “Excision of group II introns as circles”, *Molecular Cell*, Vol. 8 No. 1, pp. 201–211.

Nagy, V., Pirakitikulr, N., Zhou, K.I., Chillon, I., Luo, J. and Pyle, A.M. (2013), “Predicted group II intron lineages E and F comprise catalytically active ribozymes”, *RNA*, Vol. 19 No. 9, pp. 1266–1278.

Nielsen, P.E. (1991), “Sequence-Selective Recognition of DNA by Strand Displacement with a Thymine-Substituted Polyamide”, *Science*, No. 254, pp. 1497–1500.

Peebles, C.L. (1986), “A Self-Splicing RNA Excises an lntmn Lariat”, *Cell*, No. 44, pp. 213–223.

Pellestor, F., Paulasova, P. and Hamamah, S. (2008), “Peptide nucleic acids (PNAs) as diagnostic devices for genetic and cytogenetic analysis”, *Current Pharmaceutical Design*, Vol. 14 No. 24, pp. 2439–2444.

Podar, M., Vi T., C., Pyle, A.M. and Perlman, P.S. (1998), “Group II intron splicing in vivo by first-step hydrolysis”, *Nature*, No. 391, pp. 915–918.

Pyle, A.M. (2010), “The tertiary structure of group II introns: implications for biological function and evolution”, *Critical reviews in biochemistry and molecular biology*, Vol. 45 No. 3, pp. 215–232.

Pyle, A.M. (2016), “Group II Intron Self-Splicing”, *Annual Review of Biophysics*, Vol. 45, pp. 183–205.

Pyle, A.M. and Green, J.B. (1994), "Building a Kinetic Framework for Group II Intron Ribozyme Activity. Quantitation of Interdomain Binding and Reaction Rate", *Biochemistry*, Vol. 33 No. 9, pp. 2716–2725.

Qin, P.Z. and Pyle, A.M. (1997), "Stopped-flow fluorescence spectroscopy of a group II intron ribozyme reveals that domain 1 is an independent folding unit with a requirement for specific Mg2+ ions in the tertiary structure", *Biochemistry*, Vol. 36 No. 16, pp. 4718–4730.

Qin, P.Z. and Pyle, A.M. (1999), "Antagonistic substrate binding by a group II intron ribozyme", *Journal of Molecular Biology*, Vol. 291 No. 1, pp. 15–27.

Qu, G., Kaushal, P.S., Wang, J., Shigematsu, H., Piazza, C.L., Agrawal, R.K., Belfort, M. and Wang, H.-W. (2016), "Structure of a group II intron in complex with its reverse transcriptase", *Nature Structural Molecular Biology*, Vol. 23 No. 6, pp. 549–557.

Robart, A.R., Chan, R.T., Peters, J.K., Rajashankar, K.R. and Toor, N. (2014), "Crystal structure of a eukaryotic group II intron lariat", *Nature*, Vol. 514 No. 7521, pp. 193–197.

Robertson, K.L., Yu, L., Armitage, B.A., Lopez, A.J. and Peteanu, L.A. (2006), "Fluorescent PNA probes as hybridization labels for biological RNA", *Biochemistry*, Vol. 45 No. 19, pp. 6066–6074.

Roy, R., Hohng, S. and Ha, T. (2008), "A practical guide to single-molecule FRET", *Nature Methods*, Vol. 5 No. 6, pp. 507–516.

Russell, R., Millett, I.S., Tate, M.W., Kwok, L.W., Nakatani, B., Gruner, S.M., Mochrie, S.G.J., Pande, V., Doniach, S., Herschlag, D. and Pollack, L. (2002), "Rapid compaction during RNA folding", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 99 No. 7, pp. 4266–4271.

Saldanha, R., Mohr, G., Belfort, M. and Lambowitz, A.M. (1993), "Group I and group II introns", *FASEB Journal*, Vol. 7 No. 1, pp. 15–24.

Sanborn, M.E., Connolly, B.K., Gurunathan, K. and Levitus, M. (2007), "Fluorescence properties and photophysics of the sulfoindocyanine Cy3 linked covalently to DNA", *Journal of Physical Chemistry B*, Vol. 111 No. 37, pp. 11064–11074.

Schmitz, A.G., Zelger-Paulus, S., Gasser, G. and Sigel, R.K.O. (2015), "Strategy for Internal Labeling of Large RNAs with Minimal Perturbation by Using Fluorescent PNA", *Chembiochem: A European Journal of Chemical Biology*, Vol. 16 No. 9, pp. 1302–1306.

Schuler, B. and Eaton, W.A. (2008), "Protein folding studied by single-molecule FRET", *Current Opinion in Structural Biology*, Vol. 18 No. 1, pp. 16–26.

Schuler, B. and Hofmann, H. (2013), “Single-molecule spectroscopy of protein folding dynamics—expanding scope and timescales”, *Current Opinion in Structural Biology*, Vol. 23 No. 1, pp. 36–47.

Schuler, B., Lipman, E.A., Steinbach, P.J., Kumke, M. and Eaton, W.A. (2005), “Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence”, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 102 No. 8, pp. 2754–2759.

Sigel, R.K.O., Sashital, D.G., Abramovitz, D.L., Palmer, A.G., Butcher, S.E. and Pyle, A.M. (2004), “Solution structure of domain 5 of a group II intron ribozyme reveals a new RNA motif”, *Nature Structural Molecular Biology*, Vol. 11 No. 2, pp. 187–192.

Simon, D.M., Clarke, N.A.C., McNeil, B.A., Johnson, I., Pantuso, D., Dai, L., Chai, D. and Zimmerly, S. (2008), “Group II introns in eubacteria and archaea. ORF-less introns and new varieties”, *RNA*, Vol. 14 No. 9, pp. 1704–1713.

Smith, G.J., Sosnick, T.R., Scherer, N.F. and Pan, T. (2005), “Efficient fluorescence labeling of a large RNA through oligonucleotide hybridization”, *RNA*, Vol. 11 No. 2, pp. 234–239.

Somarowthu, S., Legiewicz, M., Keating, K.S. and Pyle, A.M. (2014), “Visualizing the ai5gamma group IIB intron”, *Nucleic Acids Research*, Vol. 42 No. 3, pp. 1947–1958.

Stanley, M. (2011), “Total Internal Reflection Fluorescence (TIRF) Microscopy (TIRFM). A Basic Overview with an Emphasis on the Optics Required”, *Nature Methods*.

Steffen, F.D. (2015), “A comparative analysis of carbocyanine photophysics in the framework of nucleic acids”, Master thesis, University of Zurich, 2015.

Steffen, F.D., Sigel, R.K.O. and Börner, R. (2016), “An atomistic view on carbocyanine photophysics in the realm of RNA”, *Physical Chemistry Chemical Physics*, Vol. 18 No. 42, pp. 29045–29055.

Steiner, M. (2008), “Metal Ion Binding to Group II Intron Ribozymes: A Study from Bulk to the Single Molecule Level”, Ph.D thesis, University of Zurich, 2008.

Steiner, M., Rueda, D. and Sigel, R.K.O. (2009), “Ca2+ induces the formation of two distinct subpopulations of group II intron molecules”, *Angewandte Chemie International Edition*, Vol. 48 No. 51, pp. 9739–9742.

Su, L.J., Brenowitz, M. and Pyle, A.M. (2003), “An Alternative Route for the Folding of Large RNAs: Apparent Two-state Folding by a Group II Intron Ribozyme”, *Journal of Molecular Biology*, Vol. 334 No. 4, pp. 639–652.

Su, L.J., Waldsich, C. and Pyle, A.M. (2005), “An obligate intermediate along the slow folding pathway of a group II intron ribozyme”, *Nucleic Acids Research*, Vol. 33 No. 21, pp. 6674–6687.

Swisher, J.F., Duarte, C.M., Su, L.J. and Pyle, A.M. (2001), “Visualizing the solvent-inaccessible core of a group II intron ribozyme”, *The EMBO Journal*, Vol. 20 No. 8, pp. 2051–2061.

Swisher, J.F., Su, L.J., Brenowitz, M., Anderson, V.E. and Pyle, A.M. (2002), “Productive folding to the native state by a group II intron ribozyme”, *Journal of Molecular Biology*, Vol. 315 No. 3, pp. 297–310.

Toor, N., Hausner, G. and Zimmerly, S. (2001), “Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases”, *RNA*, Vol. 7 No. 8, pp. 1142–1152.

Toor, N., Keating, K.S., Taylor, S.D. and Pyle, A.M. (2008), “Crystal structure of a self-spliced group II intron”, *Science*, Vol. 320 No. 5872, pp. 77–82.

Torimura, M., Kurata, S., Yamada, K., Yokomaku, T., Kamagata, Y., Kanagawa, T. and Kurane, R. (2001), “Fluorescence-quenching phenomenon by photoinduced electron transfer between a fluorescent dye and a nucleotide base”, *Analytical sciences the international journal of the Japan Society for Analytical Chemistry*, Vol. 17 No. 1, pp. 155–160.

Toro, N., Molina-Sánchez, M.D. and Fernández-López, M. (2002), “Identification and characterization of bacterial class E group II introns”, *Gene*, Vol. 299 No. 1-2, pp. 245–250.

van der Veen, R., Arnberg, A.C., van der Horst, G., Bonen, L., Tabak, H.F. and Grivell, L.A. (1986), “Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro”, *Cell*, Vol. 44 No. 2, pp. 225–234.

Vukadin, I. (2016), *Self-splicing of a wild-type group II intron ribozyme revealed by fluorescent PNA labels*, University of Zurich, Switzerland.

Waldsich, C. and Pyle, A.M. (2007), “A folding control element for tertiary collapse of a group II intron ribozyme”, *Nature Structural Molecular Biology*, Vol. 14 No. 1, pp. 37–44.

Zelger-Paulus, S. (2013), “Flourescent studies about the function and architecture of RNA under native conditions”, *Forschungskredit*.

Zhao, C. and Pyle, A.M. (2017), “Structural Insights into the Mechanism of Group II Intron Splicing”, *Trends in biochemical sciences*, Vol. 42 No. 6, pp. 470–482.

Zhao, C., Rajashankar, K.R., Marcia, M. and Pyle, A.M. (2015), “Crystal structure of group II intron domain 1 reveals a template for RNA assembly”, *Nature Chemical Biology*, Vol. 11 No. 12, pp. 967–972.

Zhao, M., Steffen, F.D., Börner, R., Schaffer, M.F., Sigel, R.K.O. and Freisinger, E. (2017), “Site-specific dual-color labeling of long RNAs for single-molecule spectroscopy”, *Nucleic Acids Research*.

Zhao, R. and Rueda, D. (2009), “RNA folding dynamics by single-molecule fluorescence resonance energy transfer”, *Methods*, Vol. 49 No. 2, pp. 112–117.

Zimmerly, S. and Semper, C. (2015), “Evolution of group II introns”, *Mobile DNA*, Vol. 6, p. 7.

Zingler, N. (2014), “The kinetics of ribozyme cleavage. A tool to analyze RNA folding as a function of catalysis”, *Methods in Molecular Biology*, Vol. 1086, pp. 209–224.

Zingler, N., Solem, A. and Pyle, A.M. (2010), “Dual roles for the Mss116 cofactor during splicing of the ai5γ group II intron”, *Nucleic Acids Research*, Vol. 38 No. 19, pp. 6602–6609.

Curriculum Vitae

SUSANN ZELGER-PAULUS

born on January 02nd 1981 in Hoyerswerda, Germany
Swiss, Stans, NW
mother of a son, Simon Zelger, 13.11.2013

EDUCATION

Since 04/2012	Ph.D. student in the group of Prof. Roland K. O. Sigel, Inorganic Chemistry, University of Zurich Title: "*Combining Single-Molecule with Ensemble Fluorescence Experiments to Disentangle the Interplay of Folding and Activity of a Group II Intron Ribozyme*"
02/2010 – 11/2010	Master Thesis at the Institute of Inorganic Chemistry, University of Zurich, Prof. Greta R. Patzke Title: "*Microscopy studies of the cellular uptake of anti-HIV Polyoxometalates*"
09/2008 – 02/2011	Master-Studies in Biochemistry, Institute of Biochemistry, University of Zurich
05/2006	Certificate: Bachelor of Science (Biotechnology), FH-Lausitz, University of Applied Sciences, Senftenberg, Germany
09/2000 – 03/2006	Bachelor-Studies in Biotechnology, FH-Lausitz, University of Applied Sciences, Senftenberg, Germany
09/2004 – 03/2005	Practical part of the Bachelor thesis at the Institute of Medical Microbiology, University of Zurich, Prof. E. C. Böttger, Title*: "Cloning and Purification of ribosomal proteins of Mycobacterium smegmatis"*

09/2002 – 02/2003	Practical training in the pharmacology company elbion AG, Radebeul, Germany, Dipl.-Biol. K. Vogel, Title: *"Development and optimization of cell cAMP-Assays for substance screening"*
09/1999 – 07/2000	Diploma-Studies in Biology and German, University of Leipzig, Germany
09/1991 – 06/1999	Grammar school, Hoyerswerda, Germany School-leaving examination
09/1987 – 07/1991	Primary school, Hoyerswerda, Germany

RESEARCH EXPERIENCES

05/2013 – 05/2017	Administrative assistant for COST Action CM1105
09/2010 – 03/2012	Research Associate at the biotech company Molecular Partners, Zürich –Schlieren, Dr. Daniel Steiner Field: *"Ribosome display selections, Protein screening, expression, purification and characterization"*
07/2006 – 01/2010	Research Technician at the Institute of Inorganic Chemistry, University of Zurich, Prof. Roland K. O. Sigel Topic: *"Metal ion-RNA-Interaction"*
03/2005 – 12/2005	Research Assistant at the Institute of Medical Microbiology, University of Zurich, Prof. Erick C. Böttger
03/2003 – 07/2003	Employee at the pharmacological company elbion AG, Radebeul, Germany, Dipl.-Biol. K. Vogel

PUBLICATIONS

- *Following RNA folding by smFRET through encapsulation*
  Susan Zelger-Paulus, Mélodie C.A.S. Hadzic, Besim Fazliji, Richard Börner, Roland K.O. Sigel, in preparation, 2018

- *Folding and cleavage kinetics revealed the influence of flanking exons for a Group II intron*
  Susann Zelger-Paulus, Besim Fazliji, Richard Börner, Roland K.O. Sigel, in preparation, 2018

- *Internal Labeling Strategy of Large RNAs with Minimal Perturbation Using Fluorescent PNA*
  Anita G. Schmitz, Susann Zelger-Paulus, Gilles Gasser, Roland K. O. Sigel, *ChemBioChem*, 2015, 16, 1302.
  Selected for Inside Cover of issue 9/16 of ChemBioChem

- *RNA three-way junctions stabilised by a molecular di-iron(II) cylinder*
  Susann Paulus, Siriporn Phongtongpasuk, Joachim Schnabl, Roland K. O. Sigel, Bernhard Spingler, Michael J. Hannon, Eva Freisinger, Angew. *Chem. Int. Ed. Engl.*, 2013, 52, 11513.
  Selected as top 5% Very Important Paper (VIP) -- Front Cover of October issue Angewandte Chemie -- UZH Press release, highlighted on various scientific news platforms

- *Targeted Delivery of Polyoxometalate Nanocomposites*
  Georg Geisberger, Susann Paulus, Emina Besic Gyenge, Carolin Maake, Greta R. Patzke, *Small*, 2011, 7, 2808.

- *Synthesis, characterization and cytotoxicity of polyoxometalate/carboxymethyl chitosan nanocomposites*
  Georg Geisberger, Susann Paulus, M. Carraro, M. Bonchio, Greta R. Patzke, *Chem. Eur. J.,* 2011, 17, 4619.

- *Using In Vitro Transcription to Construct Scaffolds for One-Dimensional Arrays of Mercuric Ions*
  Silke Johannsen, Susann Paulus, Nicole Düpre, Jens Müller, Roland K. O. Sigel, *J. Inorg. Biochem.*, 2008, 102(5-6), 1141.

AWARDS

2016	CMSZH Travel Award 2016
2016	Chemistry Travel Award by SCNAT (Swiss Academy of Sciences)
2015	SNF Forschungskredit UZH Prolongation [FK-15-094]
2013	SNF Forschungskredit UZH [FK-13-095]

TEACHING

04/2016 – 01/2017	Supervision of Master Thesis, Besim Fazliji
07/2015 – 08/2015	Supervision of BUSS student, Manuela Zimmermann

03/2015 – 01/2016	Supervision of Master Thesis, Ilija Vukadin
09/2014 – 12/2014	Chemistry lab class for medical students
09/2014 – 12/2014	Chemistry lab class for veterinary students
07/2014 – 08/2014	Supervision of BUSS student, Besim Fazliji
07/2013 – 08/2013	Supervision of Biological Undergraduate Summer School (BUSS) student, Cristina Batlle Carreras

CONFERENCE CONTRIBUTIONS

- **Following RNA splicing in vitro by single-molecule FRET** (oral presentation)
  Susann Zelger-Paulus, Mélodie C.A.S. Hadzic, Richard Börner, Roland K. O. Sigel
  RNA Society Meeting, Prague, Czech Republic
- **Combining single-molecule and ensemble experiments to disentangle the interplay of RNA folding and splicing** (oral presentation)
  Susann Zelger-Paulus, Mélodie C.A.S. Hadzic, Richard Börner, Roland K. O. Sigel
  Zing Conference about Nucleic Acids, Tampa, USA, 2016.
- **Single-molecule FRET study of an encapsulated group II intron** (oral presentation)
  Susann Zelger-Paulus, Mélodie C.A.S. Hadzic, Richard Börner, Roland K. O. Sigel
  EMBO workshop, RNA structure meets function, Stockholm, Sweden, 2016.
- **Single-molecule FRET study of an encapsulated group II intron** (oral presentation)
  Susann Zelger-Paulus, Mélodie C.A.S. Hadzic, Richard Börner, Roland K. O. Sigel
  Doktorandentag, Zurich, Switzerland, 2016.
- **Peptide Shuttle System to Deliver PNAs to their Place of Action** (poster)
  Susann Zelger-Paulus, Ilija Vukadin, Philipp Anstätt, Anita G. Schmitz, Gilles Gasser, Roland K. O. Sigel
  Fall Meeting of the Swiss Chemical Society, Lausanne, Switzerland, 2015.
- **A new labelling strategy to visualize an RNA splicing process** (poster)
  Susann Zelger-Paulus, Anita G. Schmitz, Gilles Gasser, Roland K. O. Sigel
  Fall Meeting of the Swiss Chemical Society, Lausanne, Switzerland, 2015.

- **Recognition of an RNA three-way junction by a *de novo* designed drug** (poster)
  Susann Zelger-Paulus, Siriporn Phongtongpasuk, Michael J. Hannon, Roland K. O. Sigel
  12th European Biological Inorganic Chemistry Conference (EUROBIC), Zurich, Switzerland, 2014.

- **Binding Studies Of A Designed Drug To An RNA Three-Way Junction** (poster)
  Susann Zelger-Paulus[a], Siriporn Phongtongpasuk[b], Michael J. Hannon[b], Roland K. O. Sigel
  2nd International Symposium on Functional Metal Complexes that Bind to Biomolecules, Zurich, Switzerland, 2014.

- **Imaging of an RNA-splicing process by PNA labels** (poster)
  Susann Zelger-Paulus, Anita G. Schmitz, Gilles Gasser, Roland K. O. Sigel
  DNA habitats and its RNA inhabitants, Salzburg, Austria, 2014.

- **Fluorescence studies about the function and architecture of RNA under native conditions** (poster)
  Susann Paulus, Roland K. O. Sigel
  13th Conference on Methods and Applications of Fluorescence (MAF-13), Genoa, Italy, 2013.

REFEREE EXPERIECE

Nucleic Acids Research

Conference conversation on July 2014 in Salzburg

Eric Westhof: "The group II intron you are working on is a lousy intron, choose another type of RNA for your investigations …"

Druck: www.druckterminal.de
KDD Kompetenzzentrum Digital-Druck GmbH
Leopoldstr. 68, D-90439 Nürnberh